TEAM TAO

Table of Contents

Science

9 September 2005

Volume 309

Number 5741

NEW THIS WEEK:

The Power of Walking

Toward Cheaper Genomes

As the Human Brain Evolves

Nanosprings Take Form

Research
- This Week in Science
- Reviews
- Brevia
- Research Articles
- Reports
- Technical Comments

News
- News Summaries
- News of the Week
- News Focus

Commentary
- Editorial
- Letters
- Policy Forum
- Book Reviews
- Perspectives

▶ Editors' Choice
▶ NetWatch
▶ ScienceScope
▶ Random Samples
▶ New Products
▶ Science Online Contents

RESEARCH

COMMENTARY

Toward Magnetic Spintronics

At present, microelectronic technology relies on the flow and control of electron charge in lithographically fabricated circuits, but electrons also have a quantum-mechanical spin that could be exploited to create logic circuits. Many approaches to "spintronics" have focused on semiconducting materials, but others have investigated magnetic materials. **Allwood et al.** (p. 1688; see the cover) discuss a particular proposed architecture for magnetic logic circuits that harnesses the movement of magnetic domain walls, which are the boundaries between oppositely aligned magnetic regions. Basic logic functions and nonvolatility have been demonstrated with simple combinations of magnetic nanowires.

Perovskite Preservation

The primary mineral in Earth's lower mantle is perovskite (Fe,Mg)SiO_3, but the amount of perovskite will depend on the overall composition. For example, subducted oceanic crust may be predominantly perovskite. The remixing or homogenization of these regions is affected greatly by the diffusion of major cations—Fe, Mg, and Si—at the high temperatures of Earth's mantle. **Holzapfel et al.** (p. 1707, published online 28 July 2005) have now measured the interdiffusion of Fe and Mg in perovskite in experiments at high temperature and pressure. Diffusion is so slow that homogenization, even on very small scales, is not possible on time scales as long as the age of the Earth. Thus, disparate regions in Earth's mantle will be preserved unless they are mixed mechanically.

Frustrating Magnetism

Long-range magnetic ordering that develops as a ferromagnet or antiferromagnet is cooled can be suppressed by introducing defects. Theoretical work has suggested that suppression could also occur on geometrically frustrated lattices in which the pairwise coupling of the spins cannot be met, such as on a triangular lattice. **Nakatsuji et al.** (p. 1697) now report on the realization of one such bulk triangular lattice, $NiGa_2S_4$, and present evidence that antiferromagnetic order can be completely suppressed, even to the lowest temperatures. The simple lattice structure could also be used to probe other cooperative phenomena, such as quantum criticality and second-order phase transitions.

b, a, a = 0.3624 nm

Turning Its Stripes

Under conditions of vapor-solid growth at high temperatures, zinc oxide can form nanobelts, and the dipoles set up by the opposite
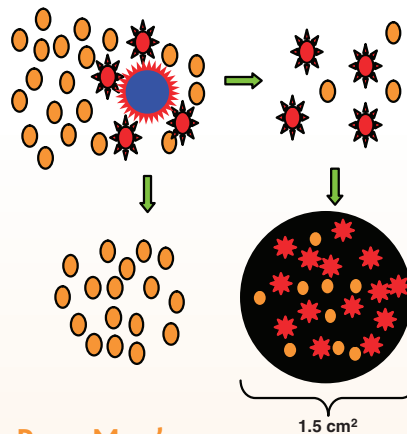
Poor Man's Sequencing?

Cheaper sequencing technologies are a high priority for many applications, including resequencing projects to study genomic variation. **Shendure et al.** (p. 1728, published online 4 August 2005; see the 5 August news story by **Pennisi**) bring the $1000 genome a step closer with a nonelectrophoretic approach, based on amplification of DNA fragments and enrichment followed by ligase-based sequencing on immobilized beads and imaging using an epifluorescence microscope. The procedure was roughly ninefold cheaper than conventional sequencing, with greater accuracy and speed. Other laboratories should be able to build their own equivalent sequencing system using off-the-shelf equipment.

1.5 cm^2

charges of the two faces of these ultrathin ribbons can drive the form of spirals or even closed loops. **Gao et al.** (p. 1700; see the Perspective by **Korgel**) now report that prolonged annealing at very high temperatures (1400°C) in an inert atmosphere can cause the belts to widen and twist into long helices. High-resolution transmission electron microscopy shows that the nanobelts are converted into a superlattice of alternating stripes that run the length of the belt and that orient their c axes at right angles to one another. This change reduces the polarity difference between the two faces, and a small twist between adjacent stripes drives the formation of the helix.

More Makes More

The effect of increasing CO_2 levels on forest soils has been studied by **Heath et al.** (p. 1711), who tracked the movement of carbon using stable carbon isotopes. High concentrations of CO_2 stimulated increased photosynthesis and plant growth but also caused a decline in the amount of carbon sequestered in the soil that was not affected by addition of soil nutrients. Microbial respiration in soils could provide a large positive feedback on the rate of increase of the concentration of atmospheric CO_2, and raise the possibility that the future rise in atmospheric CO_2 concentrations could be higher than expected.

Puzzling Brain Genes

Mutations in the genes *Microcephalin* and *ASPM* (*abnormal spindle-like microcephaly associated*) in humans correlate with microcephaly. Although the affected brains are much reduced in size, the detailed architecture of the brain remains. A gene related to *Microcephalin* is under positive selection pressures in the evolution of primate lineages leading up to humans. **Evans et al.** (p. 1717) and **Mekel-Bobrov et al.** (p. 1720) have now analyzed the evolution of *Microcephalin* and *ASPM* genes in modern humans and find that both genes also respond as though under positive selection (see the news story by **Balter**). Thus, unknown advantages have encouraged the rapid spread of these gene variants throughout the human population.

The Message Is the Messenger

As the days lengthen in spring and summer, plants sense the hours of daylight in their leaves and respond by initiating flowering at the top of the plant, in the shoot apex. The identity of the signal that is transported from the leaf to the apex has been unclear (see **Wigge** *et al.*, **Abe** *et al.*, and the Perspective by **Blázquez** in the 12 August issue). **Huang** *et al.* (p. 1694, published online 11 August 2005) now show that local expression of the gene *FT* (*FLOWERING LOCUS T*) in a single leaf is sufficient to cause flowering by transfecting a heat-inducible form of FT into *Arabidopsis* plants. Within 6 hours of stimulating a single leaf, FT messenger RNA (mRNA) appears in the shoot apex, where it stimulates transcription of genes involved in flowering and of FT itself. Although other elements may be involved, FT mRNA is an important component of the floral stimulus that moves from leaf to shoot in response to increases in day length.
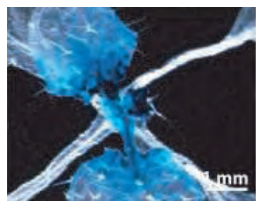
Patchwork Plant Life

The fynbos Mediterranean shrubland of the Cape Floristic Region of South Africa is one of the most species-rich plant habitats on Earth. **Latimer** *et al.* (p. 1722) find that the region's high plant diversity is associated with low migration rates and with speciation rates higher than those in the most diverse tropical forests such as Amazon rain forest. The local species abundance patterns in fynbos is a mosaic made up of centers of locally endemic plants, with little migration among them and relatively few universally rare species, unlike the well-mixed tree communities of Amazon upland rain forest.

Power Walking

Even highly rational people can be seduced by the prospect of getting something for nothing. **Rome** *et al.* (p. 1725; see the Perspective by **Kuo**) have devised a machine that recovers energy that is otherwise wasted. They have modified a backpack by introducing a vertically moveable weight that rises about 5 centimeters with each step and then turns a gear as it falls. This device can be used to recover some of the energy used in carrying supplies—a load of 38 kilograms produces up to 7 watts of electricity, compared with about 20 milliwatts from shoe-based devices. Although many refinements must be made, such an apparatus could generate power during journeys beyond the reach of power grids, reducing the need for heavy batteries.

Freeing p53

The tumor suppressor protein p53 functions to promote cell death or apoptosis in response to stress. It acts by modulating gene expression in the nucleus and by interacting with regulatory proteins in the cytoplasm that control apoptosis. **Chipuk** *et al.* (p. 1732; see the Perspective by **Vousden**) provide evidence for a mechanism by which these actions of p53 may be coordinated. The product of one p53 target gene is a protein known as PUMA (p53 up-regulated modulator of apoptosis). In cells exposed to DNA-damaging agents, interaction of PUMA with the antiapoptotic protein Bcl-xL appears to cause release of p53 that was previously bound to Bcl-xL. The released p53 may then be free to activate the cytoplasmic events that lead to apoptosis.

Plague Targets Its Victims

So-called type III secretion machines are used by *Yersinia* and other Gram-negative bacteria to inject proteins directly into target cells. The injected effector substrates are essential factors for the pathogenesis of infectious disease. **Marketon** *et al.* (p. 1739, published online 28 July 2005) now show that *Yersinia pestis*, the agent responsible for plague, selects macrophages, dendritic cells, and neutrophils for type III injection; B and T lymphocytes are rarely selected as targets. During plague, type III injection leads to the rapid depletion of immune cells from the spleen with a concomitant increase in the relative amount of injected cells. The selection of host cells with innate immune functions disables the immune system and leads to rapid progression of this invariably fatal illness.

Responsibility Beyond 60 Years

This year marks the 60th anniversary of the atomic bombings of Hiroshima and Nagasaki that ended World War II. Whatever the historical circumstances at the time, the bombings were tragic experiences for the cities and for the many who sustained severe blast, heat, and radiation injuries. The horrifying consequences prompted efforts to control nuclear arms proliferation, an issue that still remains a top foreign policy priority. The cities have made remarkable recoveries, and the survivors maintain indomitable spirits, allowing themselves to serve as a reference point for modern radiation safety in the world. As the size of this population of survivors grows smaller, we must not lose a sense of responsibility to support them and to investigate and understand the health consequences engendered by these events.

This week, the World Health Organization and Nagasaki University hold a joint meeting to review the latest information on radiation-induced effects in the atomic bomb survivors and to consider the implications of this research for international policies. Since the bombings, follow-up investigation of the health of the survivors has continued. Early in 1947, the U.S. government founded the Atomic Bomb Casualty Commission, managed by the U.S. National Academies with cooperation from the Japan National Institute of Health. In 1975, the United States and Japan became equal funding partners, creating the Radiation Effects Research Foundation to continue the study.

The main epidemiological investigation of atomic-bomb survivors is the Life Span Study. It is based on a large cohort from a general population of both sexes and all ages, encompassing a wide range of accurately known doses and incorporating accurate recordings of disease incidence and mortality. The original cohort had 120,000 survivors. This included large representative groups of young people, who are the people currently being monitored. At present, about 43% of the Life Span Study cohort is still alive.

A major finding is that radiation-associated cancers continue to occur as a small enhancement of the background rates. Through the year 2000, 8% of the deaths from solid cancers in individuals exposed to radiation could be attributed to radiation. Many of these radiation-related cancer deaths were recorded in recent years of monitoring, indicating that as the survivors age and enter the cancer-prevalent period in their lives, significant additional radiation-related deaths can be expected. Radiation-associated deaths from leukemia (46% of all leukemia deaths) peaked within 10 years of the bombings and no longer occur. Noncancer deaths are also showing a small but significant relationship, with radiation exposure making up 1.4% of these deaths. That finding requires explanation of the underlying mechanisms. Genetic effects are being investigated, although health effects in children of the survivors have not been apparent. About 94% of the second generation is still alive.

It is essential to continue the lifetime study for its full duration, for which continued cooperation from the survivors and support from the U.S. and Japanese government sponsors will be necessary. Its investigations of health consequences are heavily relied on to establish guidelines for human health and safety policies throughout the world. The survivor data feature prominently in international scientific reviews of radiation effects, such as those of the United Nations. The International Commission on Radiological Protection uses the risk evaluations to derive recommendations for radiation protection that are adopted by most countries. Reports on the Biological Effects of Ionizing Radiation from the U.S. National Academies also rely heavily on survivor data and have not only figured into formulating safety guidelines but are relevant to many current issues of concern such as worker and public protection, worker and survivor compensation, cleanup efforts, accident response, and counterterrorism preparedness.

A further 5-year funding agreement between Japan and the United States will be put in place this year, ending concerns regarding continued funding of this unique project. But a very important phase of examining the younger exposed cohorts and children of the survivors is just beginning, with a dominant time frame being the 2000–2040 period. We have a responsibility to continue this study, to contribute to the welfare of those affected, to understand and quantify the effects, and to provide a scientific basis for radiation protection and medical policies worldwide. We should not lose the memory of these events or lessen our resolve to understand their effects even after 60 years.

Burton G. Bennett

Burton G. Bennett is retiring chairman of the Radiation Effects Research Foundation, Hiroshima and Nagasaki, Japan.

EDITORS' CHOICE

edited by Gilbert Chin

CHEMISTRY
Move Away or Stay Put

Microcontact printing (µCP) of thiols on gold surfaces makes it feasible to create intricate patterns quickly and at low cost, but one limitation of this method in many applications is that the thiol molecules tend to diffuse into the bare regions of the gold surface. Trying to restrict the spreading by backfilling the bare spots in a follow-up step can blur pattern boundaries through prolonged exposure to solvent.

Dameron *et al.* show that the preadsorption of 1-adamantanethiol (AD) on gold surfaces creates a well-ordered but weakly bound layer that can be displaced by long-chain thiols, such as 1-decanethiol, that are transferred to the surface with a µCP stamp. By tuning the concentration of alkanethiols on the stamp and the contact time, patterns can be formed without the problem of lateral mixing of the adjacent self-assembled monolayers (SAMs). — PDS

Displacement printing with an alkanethiol-inked stamp.

Nano Lett. 10.1021/nl050981j (2005).

MOLECULAR BIOLOGY
Me Instead of U

Small interfering (si)RNAs and micro (mi)RNAs play a central role in RNA silencing phenomena in which RNAs complementary to the si/miRNA are targeted for cleavage and destruction, translational silencing, or are (presumably) used as markers for directing heterochromatin formation. Much is known about the biogenesis of si/miRNAs, but what happens to them afterward?

In *Arabidopsis,* the HEN1 protein, implicated in RNA silencing, has been shown to add a methyl group to the 3′ ends of miRNAs. Results from Li *et al.* show that Hen1 has a much wider scope, methylating many *Arabidopsis* siRNAs at their 3′ ends, too. In the absence of HEN1, the unmethylated miRNAs and siRNAs have several uridine (U) residues added to their 3′ ends, indicating that methylation interferes with uridylation. Intriguingly, it is the sense strand of the miRNA:miRNA* duplex that preferentially has a U tail added, hinting that the addition might occur after passage through the RISC complex, when the antisense miRNA* strand has already been jettisoned. The authors propose that methylation may stabilize small RNAs, a speculation supported by the previous observation that cleaved mRNA targets have also been observed to be U-tailed. — GR

Curr. Biol. **15**, 1501 (2005).

BIOMEDICINE
Mergers and Acquisitions

Many people living with diabetes for an extended time eventually develop a complication called diabetic neuropathy, a form of nerve damage that affects the extremities with symptoms that range from tingling and numbness to severe pain. Although many hypotheses have been proposed to explain how this nerve damage arises, the underlying mechanisms remain poorly understood.

A new study identifies an unexpected cellular culprit. From an analysis of rodent models, Terashima *et al.* find that diabetic neuropathy is accompanied by an aberrant fusion of bone marrow–derived (BMD) cells to neurons in the sciatic nerve and dorsal root ganglion. The fusion, which appears to involve a distinct subset of BMD cells that are marked by inappropriate expression of proinsulin, disrupts calcium handling in the neurons and triggers their premature death. The authors speculate that the diabetic state likely induces inappropriate gene expression in the BMD cells, producing molecules (such as tumor necrosis factor–α) that may be toxic to neurons and to the hybrid cells. Although BMD cells are often viewed in a favorable light for their potential therapeutic applications in repairing tissue damage, this study suggests that in certain contexts these cells may in fact be harmful. — PAK

Proc. Natl. Acad. Sci. U.S.A. **102**, 12525 (2005).

Hybrid cells expressing BMD and neuronal markers.

CLIMATE SCIENCE
Poring over the Past

The concentration of CO_2 in the atmosphere is understood to be a primary controller of climate, and high values are thought to be the main cause of a number of extended warm periods over the geological history of Earth. Additionally, atmospheric CO_2 exerts a fundamental influence on the carbonate chemistry of the ocean, due to the solubility of CO_2 in seawater, and therefore on calcifying marine organisms. A direct record of atmospheric CO_2 concentration is available for less than the past 1 million years, so earlier periods must be studied by modeling or via proxies.

Haworth *et al.* use measurements of the density of stomata, the pores through which gas exchange occurs in plant leaves, in an extinct conifer to estimate the partial pressure of atmospheric CO_2 during the mid-Cretaceous. This technique takes advantage of the observation that stomatal density is negatively correlated to the concentration of CO_2 in the air. They find that CO_2 was between 560 and 1200 parts per million over that period, values that fall on the low side of the range of previous estimates based on other techniques and are compatible with other evidence for relatively cool mid-Cretaceous climates. — HJS

Geology **33**, 749 (2005).

BIOCHEMISTRY
Some Like It Hot

Ever since atomic-resolution models of enzymes from thermophilic organisms appeared a decade ago, one question has been how these macromolecules are able to function (that is, remain flexible) while maintaining their integrity (that is, remain stable) at temperatures approaching 100°C. Some of the explanations offered are an increase (relative to their mesophilic cousins) in the number of salt bridges and/or hydrogen bonds, a tighter packing of the hydrophobic core, and a higher percentage of amino acids incorporated into α helices and β sheets.

Berezovsky and Shakhnovich have carried out unfolding simulations on matched proteins from mesophiles and thermophiles, and performed comparative genome-based analysis of mesophilic and thermophilic bacteria and archaea. They find two solutions for thriving at high temperatures: make protein structure more compact by optimizing relatively weak interactions globally or engineer a few strong interactions into the sequence. They suggest that archaeal thermophiles (*Pyrococcus furiosus*) were favored by starting off long ago with more designable proteins that could be adapted to the primordial hothouse, whereas bacterial thermophiles (*Thermotoga maritima*) that entered hot environments later on were forced to reinforce their proteins with staples (salt bridges or perhaps disulfides). — GJC

Proc. Natl. Acad. Sci. U.S.A. **102**, 12742 (2005).

A home for thermophiles; Vulcano Island, Italy.

CHEMISTRY
Oxides as Reductants

High-valent metal-oxo compounds are often used to mediate or catalyze oxidations of organic substrates. Nolin *et al.* have pursued the counterintuitive approach of using a rhenium-oxo complex to catalyze reduction. Rather than transferring oxygen to a hydrocarbon, the Re=O group facilitates hydride transfer from a silane to an imine. Unlike most catalysts used for this type of reaction, the Re complex is already in a high oxidation state and therefore fully stable in the open atmosphere. By appending a chiral bis(oxazoline) derivative to the Re center, the authors achieved enantioselective reductions of a broad range of aromatic imines under air at room temperature, with yields of 50 to 90% and very high enantiomeric excesses (92 to >99%). The imine nitrogens were protected with phosphinyl groups that could then be removed hydrolytically after reduction. The catalyst selectively reacts with the imine group even in the presence of esters and olefins, leading to a variety of chiral amines of potential use as pharmaceutical precursors. — JSY

J. Am. Chem. Soc. 10.1021/ja050831a (2005).

HIGHLIGHTED IN *SCIENCE*'S SIGNAL TRANSDUCTION KNOWLEDGE ENVIRONMENT

Science's STKE
www.stke.org

Neuropeptide-Mediated Receptor Trafficking

Opioid receptors on pain-sensing neurons mediate the inhibitory effects of opiates on pain. The δ-opioid receptor, which is sorted into large dense-core vesicles (LDCVs) that carry secreted neuropeptides, is inserted into the membrane in response to opioid agonists or neuronal firing. Guan *et al.* noted that, in dorsal root ganglion neurons containing substance P, δ-opioid receptors colocalized with the neurotransmitter substance P in LDCVs, but in mice lacking preprotachykinin A gene, which encodes substance P and other tachykinin peptides, the receptors were absent from the vesicles. By expressing different portions of the substance P precursor, the authors determined that the δ-opioid receptor sorting signal was in the substance P domain and that sorting into LDCVs depended on the interaction of the signal with the third extracellular loop of the opioid receptor. Stimulus-dependent membrane insertion of the δ-opioid receptor was attenuated in preprotachykinin A–knockout mice, and both δ-opioid receptor–mediated spinal analgesia and morphine tolerance were eliminated. Thus, these results suggest an intriguing link between pain pathways (substance P) and analgesia (opioid receptor). — EMA

Cell **122**, 619 (2005).

NETWATCH
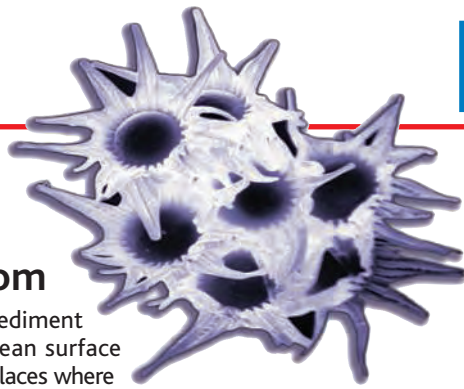
IMAGES

Microbes in Bloom

A briny desert lake and sediment 5000 meters below the ocean surface are just two of the unlikely places where microbes prosper. The Microbiological Garden, tended by Heribert Cypionka of the University of Oldenburg in Germany, shows off the bugs dwelling in these exotic environments and in habitats closer to home. The site features more than 20 photo essays on microbial topics. You can tag along on bug-hunting expeditions, learn how to isolate luminescent bacteria from herring, and observe the bugs that inhabit the scum on the surface of a stagnant pool. Some microbes make the gallery because of their beauty, such as these yeast spores (*Emericella stellamaris*; above) that resemble flowers.

www.microbiological-garden.net

DATABASE

Proteomics Central

Sequencing genomes is a breeze compared with proteomics, identifying and describing the welter of proteins that a cell, tissue, or organism harbors. PRoteomics IDEntifications database (PRIDE), a new site hosted by the European Bioinformatics Institute, helps proteomics researchers by serving as a central storehouse for experimental results. So far, the site holds data from more than 1600 studies, including fresh findings from the Human Proteome Organization's survey of the proteins in human plasma and in platelets (*Science*, 21 November 2003, p. 1316). Search the clearinghouse by species and tissue to find out which proteins and peptides each study nabbed. You can also deposit your findings in the database, which accepts results from liquid chromatography/mass spectrometry and gels.

www.ebi.ac.uk/pride/

EDUCATION

A Universal Primer

By opening Windows to the Universe, students can compare comets that frequent the solar system, study a climate change tutorial, or learn about a star's life. The encyclopedia of earth and space science from the University Corporation for Atmospheric Research in Boulder, Colorado, furnishes material from elementary school to high school levels, although many of the pages would be suitable for beginning college classes. The site's many galleries range from famous scientists to solar system objects such as the asteroid Ida (above), which is big enough to boast its own moon. Windows to the Universe also lets readers explore the interplay between science and culture by browsing poems, myths, and art about space and Earth.

www.windows.ucar.edu/windows.html

TOOLS

Physics Blog Trail

Cyber scribes could be praising or blasting your latest paper, but how would you know? If you post your work on the physics preprint server arXiv, now you can find out what colleagues are saying about it. The archive has begun displaying "trackbacks," notifications bloggers often send out when they mention a paper or site. The trackbacks appear as links below the paper's abstract, allowing anyone to see who wrote what about the findings. Read more about the feature at arxiv.org/help/trackback.

arxiv.org

EXHIBITS

Catching Rays

Tracking the seasons was so important for the Maya of Central America and many other ancient cultures that they designed buildings and settlements around the sun's annual movements. Visit some of the sun cities of North America at Traditions of the Sun, sponsored by NASA. The site's historic and modern photos, time-lapse series, video, and other multimedia whisk you off to Chaco Culture National Historic Park in New Mexico. Watch sunlight creep over the great house of Pueblo Bonito on the summer solstice, or take a virtual reality tour of the kiva at Casa Rinconada, which aligns with the points of the compass. From there you can zoom to Mayan cities on Mexico's Yucatan Peninsula such as Chichén Itzá and Dzibilchaltún. The House of the Seven Dolls (above) in Dzibilchaltún provided a spectacular visual effect for a key Mayan ceremony. On the spring and fall equinoxes, the sun rises directly behind the building and shines through the archway.

www.traditionsofthesun.org

Send site suggestions to netwatch@aaas.org. Archive: www.sciencemag.org/netwatch

NEWS

This Week

PAGE 1661
Japan's next-generation facilities

1662
The evolving brain
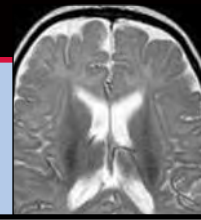
HURRICANE KATRINA

Scientists' Fears Come True as Hurricane Floods New Orleans

There are times when scientists would prefer to be wrong. Such was the case last week as Ivor van Heerden and other researchers reflected upon the devastation that Hurricane Katrina wrought on New Orleans and the Gulf Coast towns to the east. As director of Louisiana State University's Center for Public Health Impacts of Hurricanes, Van Heerden has since 2002 led a multidisciplinary team looking at what would happen if a major hurricane directly hit New Orleans. The center has studied everything from how the city would flood to how many people might ignore evacuation orders or be unable to flee—almost 1 in 4, they had estimated. "The sad part is that we called this 100%," says Van Heerden.

act upon them," says Rick Leuttich of the University of North Carolina, Chapel Hill, who has helped model how a hurricane could flood New Orleans. "We've had plenty of knowledge to know this was a disaster waiting to happen."

In one sense, Katrina, which left many researchers without homes and laboratories (see sidebar, p. 1657), was a rarity: Few hurricanes that powerful have struck the Gulf Coast in recorded history. At the same time, say hurricane experts, the storm contained few surprises. After speeding across south Florida as a category 1

Gulf of Mexico–Caribbean region." Two factors, says Olander's colleague James Kossin, fueled Katrina's growth: "phenomenally warm" waters in the gulf and a lack of strong, high-altitude winds that could have dispersed the storm's energy.

On Sunday morning, 28 August, thousands in New Orleans failed to pay heed to an evacuation order or couldn't leave. Although that shocked many, Van Heerden's center had recently polled 1000 randomly chosen New Orleans residents, using social workers to reach poor people, and had found that 21.4% would stay despite an order to leave, many of them because they lacked the means to escape.

Just before landfall, Katrina took a jog to the east, sparing New Orleans from the full force of

Katrina's wrath. These satellite pictures of New Orleans taken before (*left*) and after (*right and inset*) Hurricane Katrina give a sense of the flooding caused by breaks in the levees holding back Lake Pontchartrain in the north and the Mississippi River.

Causing the largest natural disaster in U.S. history, Katrina slammed into the Gulf Coast on 29 August with its eye hitting about 55 km east of the city. Although the storm initially brought more destruction to other areas along the Mississippi and Louisiana coast, several levees protecting New Orleans failed the following the day, and the city, about 80% of which is below sea level, filled with water. The floods may have killed thousands, stranded many more, and triggered a massive relief and evacuation effort.

Numerous studies had warned of this catastrophic scenario, and as it played out, many scientists watched with anger and frustration. "It's easy to do studies. Sometimes it's hard to

hurricane, it reached the Gulf of Mexico and began converting energy from the warm, moist air into increased intensity. By Saturday, 27 August, Katrina was a category 3 storm—and still growing.

Timothy Olander, a tropical cyclone researcher at the University of Wisconsin, Madison, recalls waking up the next morning to see that Katrina's central air pressure had dropped from about 960 millibars to below 905 millibars. The storm was now a category 5 hurricane with winds topping 175 mph. "I thought, 'Holy cow. That's an amazing development.' You don't see that rapid intensification very often," he says. Katrina "became one of the strongest storms ever recorded in the

the storm. Because of the way spinning storms interact with land, "hurricanes often wobble to the right as they come ashore," says meteorologist Hugh Willoughby of Florida International University (FIU) in Miami.

By landfall, Katrina had also shrunk to a category 4 storm. Scientists have a poor understanding of what regulates hurricane intensity, but Kossin and Willoughby note that some data indicate Katrina weakened because it had just undergone a phenomenon called eyewall replacement. The eyewall is the band of intense wind and clouds that forms around the hurricane's eye. Large storms sometimes develop an outer eyewall that starves the inner one of energy until it degrades. ▶

CREDITS: JEFF SCHMALTZ, MODIS LAND RAPID RESPONSE TEAM/GSFC/NASA; (INSET) NOAA
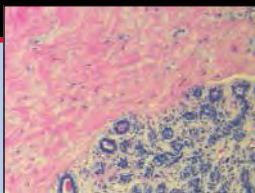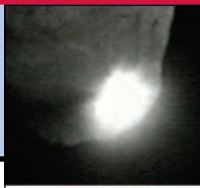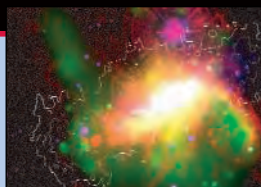
Katrina's wobble and weakening seemed at first to prevent what many have called the New Orleans "nightmare scenario." The city's main threat from hurricanes is a storm surge, the wall of water pushed onto land as the hurricane comes ashore. This surge can rise 8 meters or more as the water goes from deep water into shallow areas and then onto land. Because Atlantic hurricanes spin counterclockwise, the surge tends to be highest on their east side as the winds help any water moving north.

Because much of the city is below sea level, New Orleans is particularly vulnerable to a storm surge moving through the gulf and into Lake Pontchartrain. Over the past few decades, several computer models have shown how strong hurricanes on the right track could cause massive "overtopping" of the levees that, averaging almost 5 meters high, keep the lake from the city. The National Oceanic and Atmospheric Administration's (NOAA's) official storm surge model SLOSH (Sea, Lake, and Overland Surges from Hurricanes) was developed in the late 1960s, and Leuttich and several
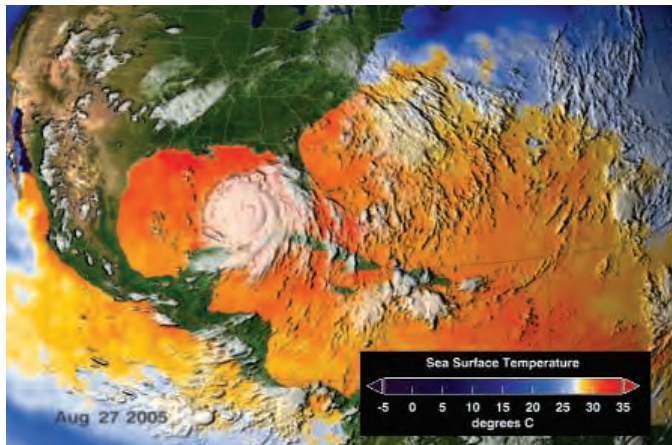
In hot water. As Katrina traveled through the Gulf of Mexico, unusually warm waters strengthened it into a monster hurricane.

collaborators have created a more sophisticated model called ADCIRC (Advanced Circulation) that has been adopted by the Army Corps of Engineers and other groups. Last year, in an exercise simulating a direct hit by a slow-moving category 3 hurricane, both models showed that the levees would not prevent the flooding of New Orleans.

According to these models, Katrina's storm surge should not have submerged the city.

Joannes Westerink of the University of Notre Dame in Indiana, who helped develop ADCIRC, says it estimated that the southern shores of Lake Pontchartrain only rose about 3 meters during Katrina. (The various models estimate that the Mississippi coast received a peak storm surge of about 7 to 9 meters, which would be the highest in U.S. history.)

Instead of overtopping, the catastrophic collapse of several levees—ones that had been upgraded with a thick concrete wall—apparently sealed the city's fate. Stephen Leatherman, director of FIU's hurricane research center, suggests that the lake's raised levels may have increased water pressure to the point that water flowed through the earthen levees on which the concrete walls sat. "Then the whole thing collapses. This is how an earthen dam collapses during a flood," he says.

The devastation from Katrina may reignite interest in bolstering the wetlands south of New Orleans to provide more of a hurricane barrier. As a storm passes over, wetlands and barrier islands along the ▶

Riding Out the Storm

Immunologist Seth Pincus survived Hurricane Katrina, but much of his research may not. Evacuated from Louisiana State University Children's Hospital in New Orleans on Thursday, Pincus left hundreds of fragile blood and tissue samples—representing years of HIV and other infectious disease research—to an uncertain fate.

Pincus, 57, studies the interaction of antibodies and pathogens and directs the hospital's Research Institute for Children. Throughout the storm, he and several hundred other hospital employees stayed to look after 100 remaining patients, as well as research samples belonging to him and colleagues. "We probably held out the longest," Pincus says. "A lot of people in New Orleans wound up abandoning their work. I think every scientist there was worried about what's more important—my experiments or my life."

The low point came 2 days after the hurricane, Pincus says. The staff realized that the lack of clean water, combined with fears of looters, posed a health risk that would force them to abandon the hospital—and the hundreds of research mice and rats that they had managed to save. Rather than let the animals starve, dehydrate, or overheat, Pincus euthanized them with pentobarbital. Then he packed what he could into insulated containers, hoping to keep cell lines and microbial collections cold until they could be transported to Baton Rouge. "Everything I own and do is [normally] in the –80°C freezer and liquid-nitrogen tanks," Pincus says.

In the end, the staff didn't want to wait for the planned afternoon exit convoy and began to leave hours ahead of schedule. "It was so hectic and crazy," Pincus says. "We had to leave probably the most important specimens." Samples packed for the trip, but abandoned, may last for a week, he says. The freezer was still running on generator power when Pincus left—but will automatically shut off unless the New Orleans SWAT team using the building as a command center keeps it running.

Pincus plans to settle in at a temporary base for the Children's Hospital set up in Baton Rouge. Although the National Institutes of Health has extended grant deadlines for flood victims, he wonders how New Orleans researchers will stay competitive, with delays of months and the loss of research samples and animal colonies. Some colleagues, he says, may choose to go elsewhere. "That's the big concern for New Orleans: If we can't get back up and going within 2 to 3 months, anyone who can go anywhere else will."

Pincus tries to remain hopeful. "I may have to start all over again," he says. "But maybe this is an opportunity to take some novel approaches. In some ways, it may even be liberating."

–CAROLYN GRAMLING

coast sap its energy and reduce storm surge. By some estimates, however, up to 100 square kilometers of this buffer disappear each year, largely because the Mississippi River has been leveed and dammed so much that it deposits much less sediment onto the delta. (Katrina wiped out barrier islands herself.)

In 1998, a collection of state and federal agencies, including the Environmental Protection Agency and the Army Corps of Engineers, proposed Coast 2050, a $14 billion strategy to restore Louisiana's wetlands (*Science*, 15 September 2000, p. 1860). But the project never won federal funding and hasn't moved beyond the planning stage.

Any renewed debate over coastal restoration will likely have to await the long cleanup and recovery of New Orleans and the surrounding areas, a process that will take months, if not years. Among the research institutions bearing the brunt of Katrina were Tulane University, the University of New Orleans, and Xavier, all of which lost power to most of their campuses. The three universities have canceled their fall semesters and are still surveying damages.

The National Institutes of Health (NIH) has begun working with the medical schools of Louisiana State University and Tulane in New Orleans, which together have about 300 NIH grants totaling $130 million annually. NIH is trying to arrange for temporary lab space at other universities for research teams displaced by the flooding. "It's clear that the time to return to facilities that will be functional is undefined at this point," said NIH Director Elias Zerhouni. "We need to make provisions to continue their research and be able to support them during this interim period." NIH was also working to get generators and other support to help New Orleans–area researchers keep lab animals alive and samples cold.

Farther from the floodwaters, the picture is less bleak. The deluge from Lake Pontchartrain never reached Tulane's National Primate Research Center (TNPRC), situated on higher ground on the opposite side of the lake. Still, Katrina's winds caused extensive damage to the facilities, says Tom Gordon of Yerkes National Primate Research Center in Atlanta, Georgia, who was contacted by one of the TNPRC employees who stayed behind to care for the animals. "In the last days, [the center] took some extraordinary precautionary measures to do what they could," Gordon says.

Also weathering the storm were several major research facilities, including the Laser Interferometer Gravitational Wave Observatory at Livingston, Louisiana, and NASA's Stennis Space Center in Mississippi. Lockheed Martin's Michoud Assembly Facility in New Orleans, which builds the external fuel tanks for the space shuttle, suffered no injuries and only minor wind

CREDIT: J. WESTERINK/UNIVERSITY OF NOTRE DAME
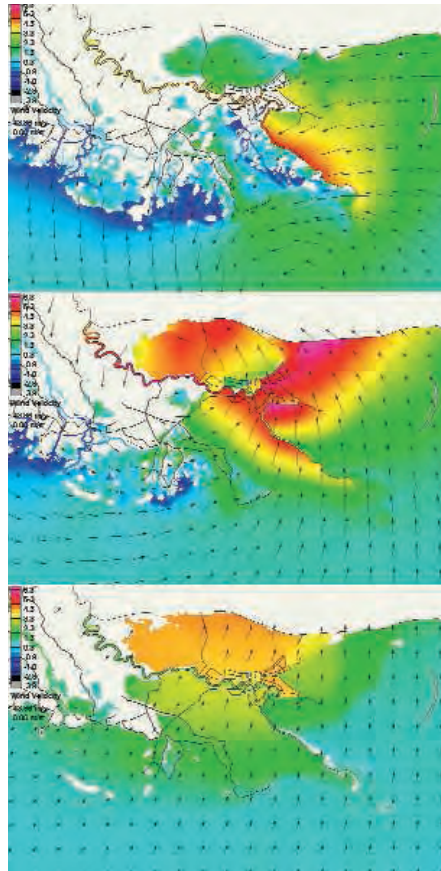
Deadly surge. A computer model of a fictional category 3 hurricane shows how a storm surge moves from the gulf into Lake Pontchartrain and floods the New Orleans area.

and roof damage, says June Malone, a spokesperson for NASA's Marshall Space Flight Center in Huntsville, Alabama.

The gulf itself may sustain damage from Katrina. Data are still coming in to NOAA, but two large oil spills from coastal storage facilities have been identified. "We would expect environmental impacts," says Tom Callahan of NOAA's Hazardous Materials Response division.

Katrina scuttled at least one major scientific meeting. The Interscience Conference on Antimicrobial Agents and Chemotherapy was scheduled for New Orleans later this month but will now take place in Washington, D.C., in December. Ironically, New Orleans was also going to host two major hurricane research conferences in the spring.

Even as the New Orleans region and its research institutions struggle to recover from Katrina, many are casting nervous eyes to the gulf. Hurricane season is far from over this year, and researchers say that the United States has entered a period that is likely to bring more major hurricane strikes. "That's scary," says Olander. "Who knows what else is on the way?" **–JOHN TRAVIS**

With reporting by Carolyn Gramling, Jocelyn Kaiser, Eli Kintisch, and Erik Stokstad.

Committee: Nurture DOD Tissue Bank

An expert panel has recommended that the imperiled Armed Forces Institute of Pathology's (AFIP's) vast tissue bank stay in government hands and be given adequate resources.

AFIP in Washington, D.C., is being closed and its parts transferred as part of the latest round of base closings (*Science*, 2 September, p. 1472). A terse Department of Defense (DOD) recommendation to preserve the 3-million-case repository, however, sparked concerns that it would be mothballed without needed staff and expertise. At a conference last week in Washington, a 12-member panel chaired by former Stanford medical school dean David Korn concluded in a draft statement that the repository is in "excellent condition" and "should be maintained as a vibrant, living" resource and made more widely available to outside scientists. The U.S. government should retain ownership and provide adequate professional and technical staff, the panel said. Congress may need to enact statutory authority to carry out the panel's final recommendations, which were to go to DOD this week. **–JOCELYN KAISER**

Roberts's AIDS Memo Criticized

Supreme Court Chief Justice nominee John Roberts may find himself confronted during Senate hearings by a position he took about AIDS when he worked for President Ronald Reagan.

In a September 1985 memo about an upcoming presidential press conference, Roberts, then an associate counsel at the White House, advised Reagan not to take sides on the question of whether schoolchildren with AIDS could infect their classmates. Two weeks before Roberts wrote the memo, recommendations from what was then called the Centers for Disease Control said casual contact in schools "appears to pose no risk." But in the memo, Roberts called the question a "disputed scientific issue." Reagan apparently took his advice, saying "[M]edicine has not come forth unequivocally and said, 'This we know for a fact, that it is safe.' "

Representative Henry Waxman (D–CA) highlighted and assailed the memo in a 1 September letter to senators Arlen Specter (R–PA) and Patrick Leahy (D–VT), who will lead the Roberts hearings. "There was quite a bit of information back then on the lack of household transmission," says Gerald Friedland, head of the AIDS program at Yale University. **–JON COHEN**

PLANETARY SCIENCE

At Last, a Supportive Parent for Saturn's Youngest Ring

Imperceptible, nearly undetectable, the ephemeral E ring of Saturn shouldn't be there at all. Its icy, micrometer-size particles can't possibly survive more than a few centuries in their orbit far beyond the main rings, so something must be replenishing the E ring. For 25 years, planetary scientists have looked to Saturn's moon Enceladus for an explanation; perhaps icy volcanoes on the inscrutable satellite are the source. Last July, the Cassini spacecraft did detect volcanolike activity on the 500-kilometer satellite, only to find that no particles were being ejected. Nothing but water vapor was shooting off the newly found hot spot near the south pole (*Science*, 5 August, p. 859).

Now, with further analysis, Cassini scientists have decided that what may be geyserlike venting at the hot spot is in fact launching ice particles into the enveloping E ring, and in sufficient quantities to sustain it. "Before, we were puzzled," says Cassini team member Larry W. Esposito of the University of Colorado, Boulder. "Now we have a plausible story."

At first, July's close flyby of Enceladus seemed to support a relatively mundane source for the E ring, says Frank Spahn of the University of Potsdam, Germany, a member of Cassini's Cosmic Dust Analyzer team. As Cassini passed within 175 kilometers of the
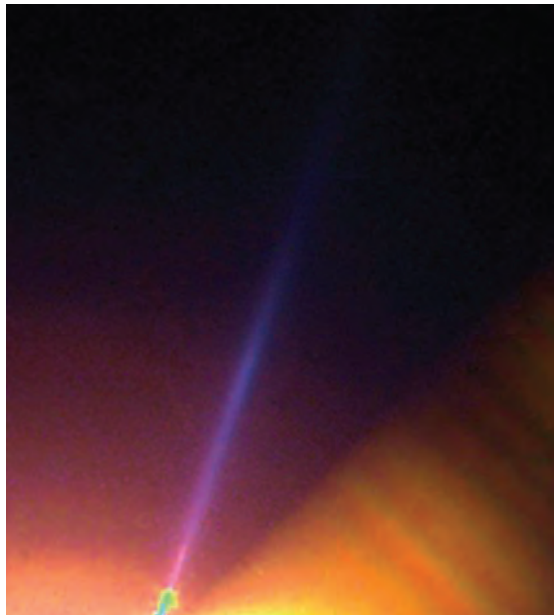
An orphan no more. Geysers on Enceladus replenish the E ring (here in false color from the Hubble Space Telescope).

moon, the number of particles detected by the dust analyzer rose and declined smoothly. That's just what computer models suggested Cassini would see if the steady rain of micrometeorites that strikes Enceladus erodes E ring particles from the entire surface of a geologically dead moon. But Cassini instruments also detected a relatively hot spot (a cool 85 K) near the south pole, where particularly warm gouges in the surface were giving off water vapor. The hot spot caught Spahn's attention, as did the way the sharp

peak in venting water came a minute later than the broad peak in dust hits. So Spahn and his colleagues went back to their models.

With further work, the Potsdam group found that only a small, localized dust source could produce a broad peak in dust detections as well as the 1-minute difference between dust and water peaks. "We're astonished, [but] there's no other way," says Spahn. The particles "have to come from the south pole."

Using Cassini ultraviolet observations of a star passing behind the south pole emissions, Esposito and his team members also find that if even a small proportion of the water vented by Enceladus is in the form of ice particles, there would be enough to sustain the E ring. There would also be enough particles to account for last year's E-ring outburst of atomic oxygen observed by Cassini (*Science*, 14 January, p. 202), he says. He now postulates that either geyserlike activity on Enceladus surged then or that Saturn's magnetosphere accelerated its charged-particle erosion of E-ring particles.

The 25-year search for the E ring's source seems to be over, but that only shifts attention to the next mystery: What is Enceladus doing with a warm south pole in the first place? It is too small and old to generate enough heat by radioactive decay. Jupiter's fiery Io heats up through gravitational interactions with other moons, but Enceladus is not in one of the usual orbital relationships with another satellite that would allow tidal energy to be pumped in. One thing's for sure, though: The ball is now in the orbital dynamicists' court.

–RICHARD A. KERR

PHYSICS

Journal Plans Faster Thumbs-Down for Rejected Papers

A good editor can quickly sift the wheat from the chaff in a pile of physics papers. At least that's what the editors of the leading physics journal, *Physical Review Letters* (*PRL*), have decided. Last month, they announced they will increase the fraction of submitted papers they reject without peer review from between 10% and 15% to between 20% and 25%. The move should ease the burden on referees, speed up the vetting process, and help *PRL* compete for papers with *Science* and *Nature*.

PRL has always summarily rejected papers that are obviously wrong, says editor Jack Sandweiss of Yale University in New Haven, Connecticut. Editors have quickly rejected another 5% to 10% of submissions because they clearly are not important and accessible to a broad range of physicists, as *PRL* papers are supposed to be. Internal

studies show, however, that editors can predict which papers referees will ultimately reject on the same grounds, Sandweiss says, so editors will now make that call more often. "We're not changing the criteria for *Physical Review Letters*," he says. "We're just applying them earlier."

The new policy should help *PRL* cope with a rising tide of manuscripts. Editors expect more than 10,000 submissions this year, twice as many as they received 12 years ago. *PRL* accepts about 35% of submissions, and the median time to acceptance is more than 4 months, says editor Stanley Brown from his office in Ridge, New York.

Physicists are cautiously optimistic about the new approach. If a paper is going to be rejected, it's better to find out sooner so it can be submitted to another journal

more quickly, says Raffi Budakian of the University of Illinois, Urbana-Champaign. Speeding up the selection process could help *PRL* vie for high-profile papers with *Science* and *Nature*, which reject most submissions without peer review and accept papers more quickly. Speed of review is "a big factor" in deciding which journal to submit a paper to, says J.C. Séamus Davis of Cornell University in Ithaca, New York: "Sometimes you need to get the results out there in a few weeks."

Authors can appeal a rejection. But an editorial on the *PRL* Web site (prl.aps.org) encourages authors to send rejected papers to more appropriate journals. In January, the *Journal of the American Chemical Society* announced a similar change in policy.

–ADRIAN CHO

Supercomputer, X-ray Source Win Backing in Japanese Budget

TOKYO—Japanese scientists hope that the government will ante up $67 million in 2006 to begin building a next-generation supercomputer and an advanced x-ray source. Last week, their hopes were raised when the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) included the projects in its annual budget submission (*Science*, 2 September, p. 1473). But they can't claim the pot until the end of a new science budget game.

Just a few years ago, the ministry's endorsement would have been sufficient to win funding. But now a revamped Council for Science and Technology Policy, nominally headed by the prime minister, joins the game with a list of government-wide priorities and proposals ranked according to those priorities as well as a project's scientific merit. And with the central government hoping to rein in public spending, the rankings are becoming even more important.

One council priority is strengthening the country's scientific infrastructure. (Another is

Making room. RIKEN's Harima Institute hopes to add a free-electron laser to its Spring-8 synchrotron.

the development of human resources, for which the ministry responded by expanding grants for young researchers.) And that's where both big projects shine. The proposed 10 petaflops (10 quadrillion calculations per second) supercomputer would be several times faster than any of today's machines. The ministry is asking for $37 million as a down payment on a $1 billion, 7-year project. "Computer simulation has become exceedingly important in manufacturing, nanotechnology, life sciences, earth sciences, astronomy, and other fields," says Noriyuki Matsukawa, director of the information division at MEXT.

There's also the matter of national pride. The Earth Simulator was the world's fastest when it debuted in 2002 (*Science*, 1 March 2002, p. 1631), a title now held by IBM's Blue Gene/L machine at Lawrence Livermore National Laboratory. "There is strong support for trying to reclaim this gold medal," says Akira Yoshikawa, director of MEXT's science and technology policy bureau.

The x-ray source will contribute to both fundamental and applied research in life sciences and nanotechnology, says Yasunori Kojima, director of the ministry's office of synchrotron radiation research. The next step beyond synchrotrons, which produce x-rays by sending electrons in a wide circle, is the so-called free-electron laser (FEL), which generates x-rays by manipulating an electron beam traveling in a straight line. Kojima says x-ray FELs promise to resolve thorny protein membrane structures and capture images of chemical reactions as they happen. The ministry is asking for $30 million to begin building the 5-year, $365 million facility at RIKEN's Harima Institute in Kobe.

Both projects are expected to keep Japan globally competitive. U.S. researchers are also laying plans for a petaflops supercomputer, and both Germany's synchrotron radiation laboratory (DESY) and Stanford University's Stanford Linear Accelerator Center (SLAC) are planning their own FELs. "If we don't get funding next year, SLAC and DESY will be advancing in this field, and Japan will be retreating in a big way," says Hideo Kitamura, a RIKEN physicist who leads one of the groups working on the x-ray FEL.

Ministry officials feel good about the funding prospects for the supercomputer and x-ray FEL projects, says Yoshikawa, who admits that other portions of the ministry's requested 9.5% increase in science-related spending, to $8.3 billion, are vulnerable. Last year, MEXT obtained a 1% increase in its science budget while science spending shrank by 2% at the Ministry of Economy, Trade, and Industry and by a whopping 22% for defense-related research. "We think there is recognition that academic research produces the seeds for industrial competitiveness," Yoshikawa says. But every year scientists have to prove it.

–DENNIS NORMILE

First Woman Head for Pasteur

PARIS—Cell biologist Alice Dautry was appointed managing director of the Pasteur Institute last week, the first woman to head the venerable but troubled lab. She replaces Philippe Kourilsky, who resigned on 31 July amid conflicts over his management style and a controversial plan to move part of the lab to a Paris suburb.

Dautry, 55, has been with Pasteur since 1977 and leads a 10-member group studying immune system receptors and intracellular bacteria. She is a "very serious and practical" researcher and manager, says colleague Brigitte Gicquel of Pasteur, qualities which should help calm Pasteur's waters. When Dautry chaired a panel drawing up an alternative for the move, "it took her 2 weeks to come up with a plan that was acceptable to everybody," Gicquel says.

–MARTIN ENSERINK

Ban on Papers Lifted

The American Institute of Aeronautics and Astronautics (AIAA) last week permanently ended a ban on papers submitted to its journals and conferences by authors living in countries under a U.S. embargo. (The countries are Iran, Cuba, North Korea, and Sudan.) The decision, passed by a unanimous vote at the society's board meeting in Long Beach, California, reverses a policy that was instituted in May to "avoid assisting a nation such as North Korea in its efforts to develop nuclear weapons and the capability to deliver them" (*Science*, 17 June, p. 1722). After protests from Iranian scientists, AIAA suspended the policy in June.

–YUDHIJIT BHATTACHARJEE

Consumers Nano-Cautious

Only 11% of Americans believe that voluntary standards among companies selling nano-based products are adequate to protect human and environmental safety, a new U.S. survey has found. The most in-depth study to date, the survey polled focus groups involving 177 consumers nationwide and was released yesterday by the Woodrow Wilson International Center for Scholars in Washington, D.C. It found that although individuals don't support a ban on nanotechnology in consumer products, most prefer to know what nanomaterials are in their medicines or ice cream and would like increased premarket testing to ensure that products are safe.

–ROBERT F. SERVICE

on these abilities," says Tyler-Smith, who calls the possible links to events in human prehistory "highly speculative."

Lahn and colleagues also found a pronounced pattern in the distribution of the favored alleles in populations around the world: The *microcephalin* allel, for example, is much more common in Europe, Asia, and the Americas than in sub-Saharan Africa. Using a larger sample from 1184 individuals, the team found this allele in roughly 75% or more of Italians, Russians, and Han Chinese, and in nearly 100% of Colombians. In contrast, the allele had frequencies of less than 10% in the

Zime of Cameroon and the San of Namibia, and about 30% of Tanzanian Masaai. The *ASPM* allele also showed a skewed geographic distribution.

Lahn and his co-workers say that several scenarios could account for the pattern. For example, the favored alleles may have arisen outside Africa, or they may reflect a genetic "bottleneck" that occurred when a relatively small founding population carrying the alleles migrated out of Africa.

The possibility that the favored alleles might confer some sort of cognitive edge— and that they are unevenly distributed in human populations—raises social and ethi-

cal issues, researchers say. Lahn warns that there is "a lot of potential for over- and misinterpretation" of his results. He points out that other advantageous alleles might have a very different population distribution. "You don't necessarily come out ahead" if you have these alleles, Lahn says: "We only picked out two."

Although they acknowledge such social concerns, most scientists who spoke to *Science* say that the only way to answer the questions posed by this research is to do more research. "We should treat these genes just like any others," says Tyler-Smith.

–MICHAEL BALTER

NUCLEAR MEDICINE
Panel Puts Eventual Chornobyl Death Toll in Thousands

VIENNA, AUSTRIA—A study released this week predicts that 4000 people or even more will die from cancers caused by the 1986 Chornobyl nuclear accident, a figure that dwarfs the 50 known deaths linked to the disaster so far. The report,* compiled by the Chernobyl Forum, a joint effort of eight United Nations agencies and the governments of Ukraine, Belarus, and Russia, also highlights the thousands who are suffering a variety of mental health problems since the accident.

The meltdown of one of the reactors at the Chornobyl power plant in Ukraine on 26 April 1986 released approximately 50 tons of radioactive material into the atmosphere, contaminating an area inhabited by 5 million people. Because the most pernicious contamination was radioactive iodine-131, which lodges in the thyroid, most of the casualties are expected to succumb to thyroid cancer, which typically takes 25 years or more to show up.

Over the 19 years since the accident, estimates of the final death toll from radiation-induced cancer have ranged from zero to tens of thousands. The panel of 100 scientists involved in the Chernobyl Forum reduced that uncertainty by reviewing all available data and discounting studies that were not sufficiently rigorous. "But that only considers the 600,000 people living in the most exposed areas. [The total] could double to 8000 if you also consider people around that area," says forum member Fred Mettler, a radiologist at the University of New Mexico in Albuquerque.

Radiation biologist Mikhail Balonov was part of the Soviet team rushed in to assess Chornobyl in 1986, and he says his

team "also predicted 4000 deaths. But our conclusions were classified." The forum's 600-page report, released by the International Atomic Energy Agency (IAEA) here on 5 September, also echoes initial predictions that the radiation will have no effect on fertility or the frequency of birth defects in the second generation. "Luckily, the exposure was too low for that," says Balonov, who now heads IAEA's Radioactive Discharges Unit. Other effects of the radiation are either too subtle or have not yet been detected.

The outlook for the environment around Chornobyl appears somewhat better.

Sleeping giant. A guard walks past the remains of Chornobyl's reactor #4, which is encased in a now-crumbling sarcophagus.

According to the report, 90% of the radioactive contamination was cleaned up through a massive removal of surface soils. Researchers are developing special salts and fertilizers to inhibit the remaining radioactive material in soil from getting into crop plants. But on the whole, the forum concludes, most of the originally exposed area is close to background levels of radiation.

The report's most surprising conclusion is that mental health problems appear to be more common than any radiation-linked disease. The incidence of high anxiety is twice normal levels, and unexplained pain or debilitation is three to four times that in similar unexposed populations. One possible cause is the trauma experienced by the 350,000 residents who were forcibly relocated.

Mettler, a member of the international scientific team that first visited the Chornobyl site in 1990, says another factor "is the psychological impact on people of not knowing the extent of contamination or the real health risks it poses." That uncertainty, according to the report, seems to have translated into unhealthy lifestyle choices such as heavy smoking, drinking, drug use, and poor diet.

Removing anxiety won't be easy, says Balonov. People in the Chornobyl area do not trust government officials, he notes, because "there was a tradition of lying" in Soviet times. Mettler hopes the Chernobyl Forum report will reassure residents. "It's a start," he says.

–JOHN BOHANNON

John Bohannon is a writer in Berlin, Germany.

Chernobyl's Legacy: Health, Environmental, and Socio-economic Impacts, www.iaea.org/ NewsCenter/Focus/Chernobyl

Many studies have shown that high breast density boosts a woman's chance of breast cancer. But it's not clear why, and doctors aren't sure how to translate the knowledge they've got to the clinic

Dissecting a Hidden Breast Cancer Risk

On a typical day, at least 150 women undergo mammography at the California Pacific Medical Center in San Francisco, and for the last year, most have agreed to take part in a giant experiment. Their mammograms are stripped of identifiers and transmitted across town to the University of California, San Francisco (UCSF). At both institutions, physicists, epidemiologists, and doctors are sifting through clues, trying to bring into the clinic one of the biggest—but relatively unheard of—risk factors for breast cancer: breast density.

Roughly 50 studies have suggested that high breast density—in other words, relatively little fat and lots of connective tissue in the breast compared to other women in the same age group—boosts a woman's risk of breast cancer, and current estimates put the increase at four to six times. Only two other traits are known to increase risk more: age, or harboring a mutated version of the breast cancer susceptibility genes *BRCA1* or *BRCA2*.

Despite the consistent connection between breast density and cancer, fundamental biological questions remain. It's not entirely clear what dense tissue is made of, why it increases the likelihood of breast cancer, or how it modulates an individual woman's cancer risk. Nor do researchers know whether decreasing a woman's density will lower her chance of cancer. Even measuring density is problematic (see sidebar, p. 1665).

"More and more people are realizing—we've got this strong risk factor, we need to understand it," says Celia Byrne, a cancer epidemiologist at Georgetown University Medical Center in Washington, D.C.

Faced with a powerful risk factor that comes with many unknowns, researchers and physicians are uncertain about when and how to apply their emerging understanding of breast density to patients. If integrated into clinical practice, density measures are likely to reshape the landscape of breast cancer risk assessment and prevention. The number of women with dense breasts is

enormous. High-risk clinics "would be overwhelmed" if they were all referred there, says Frederick Margolin, a radiologist and director of the Breast Health Center at California Pacific Medical Center. He's working with UCSF scientists on the mammogram study, which is part of the broader San Fran-

On the frontlines. Toronto epidemiologist Norman Boyd is pursuing the unsolved mysteries of breast density.

ciso Mammography Registry examing risks and outcomes of breast cancer.

But should women even be told their status, or would that provoke needless anxiety? Should they be offered extra screening, or prevention drugs, which carry their own hazards and whose effects on density are unclear? One physician estimates that if density were taken into account in risk assessment, up to 20% of postmenopausal women in the United States would be eligible for prevention drugs.

Some scientists believe it's unethical to discuss density with patients until the risk factor is more fully dissected; others believe it's unethical not to. But the reality is that today, few physicians aware of density's risks share the information with their patients. "We don't know what to do with it," concedes Margolin.

Density uncovered

The breast density field was born in the mid-1970s, when, on a hunch, a Michigan radiologist named John Wolfe began outlining pools of dense tissue with a wax pencil on mammograms. He reported that women with denser breasts were more likely to contract breast cancer. The increase in risk Wolfe postulated from high breast density was staggering, as much as 20 times the risk faced by women with low density.

"My initial reading was, this may be something Wolfe can see, but the rest of us probably can't," says Norman Boyd, a cancer epidemiologist at Princess Margaret Hospital in downtown Toronto. Dressed trimly in a blue button-down shirt and tan slacks, Boyd's British accent has faded after four decades in Canada. He says his early skepticism dissipated when he asked ten radiologists to assess density on mammograms. The radiologists agreed more on density than any other mammographic quality, such as tumors.

In the following years, Boyd and a small band of others helped establish that breast density increases the risk of breast cancer, though not as much as Wolfe predicted. The precise boost in risk, though, remains uncertain.

From the beginning, density researchers faced skepticism on several fronts. One lingering concern was whether high density—which makes mammograms tough to read because dense tissue appears as roughly the same shade as tumors—was masking cancer rather than driving it. In 1995, in the *Journal of the National Cancer Institute*, Byrne and her colleagues published a study

of 4000 women, 1880 of whom developed breast cancer, that she thought would settle the question. Examining mammograms pulled from storage, Byrne's group found that some women didn't develop cancer until 14 years after high density had been detected. "This was pretty convincing evidence" that dense breasts weren't just masking tumors, she says. Still, some radiologists remain unconvinced that density can augment cancer risk.

How, then, might dense tissue be triggering cancer? One of the few basic biologists exploring this is Thea Tlsty, a molecular pathologist at UCSF. She's collecting mastectomy specimens from healthy breasts, taken from women who died of other causes, such as car accidents.

Tlsty and others have found that dense breast tissue contains two types of connective tissue, along with epithelial cells, which can form tumors. The connective tissues include collagen, a protein in skin and bone, and fibroblasts, another common connective tissue in the body.

While epithelial cells are considered cells "at risk" in the cancer world, because they make up the bulk of most tumors, it's the connective tissues that have Tlsty intrigued. Just like tumors, dense breasts are "chock-full of collagen," Tlsty says. More mysteriously, she is finding that collagen in high-density breasts is actually different from collagen in low-density breasts, turning over much more rapidly.

Now Tlsty is trying to learn more about fibroblasts from her mastectomy samples. Although fibroblasts don't form tumors themselves, they can signal to cancers and help them grow. Tlsty wonders whether fibroblasts in dense breasts send the same kind of signals they send in the presence of a cancer.

Density DNA

Examining dense tissue, as Tlsty does, is one place to start. Another is in the genes. Three years ago, Boyd, his Australian colleague John Hopper from the University of Melbourne, and others reported in the *New England Journal of Medicine* that breast density was inherited among 950 pairs of identical and fraternal twins. They estimated that genetics accounts for roughly 60% of variations in density.

Now the hope is that the hunt for genes behind breast density might also turn up genes that confer breast cancer risk: The same genes may have both effects. "I don't think anybody's looked at breast cancer genes in this particular way," says Johanna Rommens, a geneticist at Toronto's Hospital for Sick Children. Just two genes, *BRCA1* and *BRCA2*, are relatively common and known to significantly increase the chance

Fine-Tuning Breast Density Measures

In addition to unresolved questions about how density might spur cancer, another barrier that prevents clinicians from using density data to gauge cancer risk is the difficulty in measuring it. "We all agree on how much is a pound and how much is an inch, but there isn't necessarily a common scale for density," says Celia Byrne of Georgetown University in Washington, D.C.

Like Norman Boyd, a founder of the breast density field who's based at Princess Margaret Hospital in Toronto, Byrne measures density the old-fashioned way: by "drawing," with some computer guidance, over digitized mammograms. The method is time-intensive and somewhat subjective. It has also fueled worries that despite their predictive power, density measures are crude and may vary in different hands.

Researchers in Toronto and in San Francisco are independently trying to improve and automate density measures. They have developed software that measures density volume, instead of just its surface area; both versions are now being tested to see if they're more predictive of risk.

In his office at Toronto's Sunnybrook and Women's College Health Sciences Centre, with stacks of papers balanced precariously on most surfaces, medical biophysicist Martin Yaffe demonstrates new software designed to measure density volume. On one breast, traditional 2-D measures reveal 7% density; the volumetric approach reveals 22%. Yaffe and Boyd are conducting further tests to validate the volumetric software, including in a study of 1000 postmenopausal Caucasian and Chinese women.

At the same time, researchers at the University of California, San Francisco (UCSF), and California Pacific Medical Center are trying to automate density fully. They're testing their method on tens of thousands of volunteers undergoing mammograms at California Pacific.

Shades of risk. Viewed on 3-D software, dense tissue in a high (62%) density breast glows red, while a breast with just 4% density is filled with cooler color tones.

"We don't have an automated program yet that's as good as having a person look at [a mammogram]," says Byrne. The leaders of the California Pacific study—including UCSF epidemiologists Karla Kerlikowske and Steven Cummings, and physicist John Shepherd, who designed the software—are hoping for an answer within the next year or so.   **–J.C.**

of breast cancer, but many families with the disease don't carry these mutations.

A fast-talking gene-hunter whose face lights up when she discusses DNA, Rommens helped discover the cystic fibrosis gene while a postdoctoral fellow. The photo of a smiling, gap-toothed 9-year-old girl is pasted above her desk, a survivor of a bone marrow failure disorder whose genetic culprit she also tracked down. But none of her previous searches, Rommens says, match her hunt for breast density genes in its complexity.

After a couple years of lobbying, Rommens agreed to join Boyd of Princess Margaret across the street in an ambitious breast density genetics project. The pair, along with several colleagues, has applied for $4 million in funding and is planning a

study of about 5000 individuals, most of them sister pairs who belong to a global breast cancer registry. (Not all the sisters have developed breast cancer.) The scientists will compare breast density on mammograms with their subjects' DNA.

Looking for density genes is complicated, says Rommens, because density is only partly genetic and unlike disease, it's not a yes or no proposition. One woman might have 20% density, another 50%, and another 75%. But no patient has 50% of diabetes.

At least two other teams are pursuing density genes: One led by cancer epidemiologists Celine Vachon at the Mayo Clinic in Rochester, Minnesota, and Thomas Sellers at the H. Lee Moffitt Cancer Center & Research Institute in Tampa, Florida; and the

other by Douglas Easton at the University of Cambridge, U.K. Easton, who helped discover *BRCA2*, is also completing a density study in 500 *BRCA* carriers, a group whose connection with breast density, if one exists, remains unclear.

Another looming question is how age and hormone levels influence the concentration

A Sampling of Breast Density Studies			
PURPOSE	LOCATION	# OF SUBJECTS	END DATE
Diet's effects on density and risk	Toronto, Canada	4700	2005
Test automated density measures	San Francisco, CA	35,000	2006
Identify density genes	Rochester, MN	900	2006
Test at what age density predicts risk	London, U.K.	35,000	unknown
Density* variation in young women	Toronto, Canada	250	2006
Examine biology of dense tissue	San Francisco, CA	n/a	unknown

Density is measured here as water content on an MRI

of dense tissue. Nearly all of the density data collected so far is on women over 50. But a woman's density levels naturally dip with age, and researchers are wondering more and more whether it's her density at age 30 that helps determine her cancer risk at age 60. "You need to ask yourself, 'Is it where they started [with density] or where they ended up?'" says Sellers.

It's questions like these that have some researchers particularly excited, because they could challenge classic thinking about breast cancer. Breast cancer gene studies, for example, says Easton, have tended to focus on the same pathways, the same hormonal patterns. "We need to break out of that paradigm," he says.

Already, some density experts are pushing work forward on a hormone that's been tentatively tied to breast cancer. Called insulin-like growth factor type 1, or IGF-1, it's produced in the liver and the breast's connective tissue. "IGF is the best clue" tying density and cancer together, says Steven Cummings, an internist and epidemiologist at UCSF, even if its precise connection to breast cancer is murky.

Clinical conundrums

Solving these questions, along with the measurement challenges, may open a Pandora's box for doctors and their patients. Not all radiologists are aware that density is associated with cancer, and even those who are may be reluctant to share density information with women for fear that it will provoke needless anxiety.

Jack Cuzick, a cancer epidemiologist at the Wolfson Institute of Preventive Medicine in London, is incensed that women aren't told their breast density measures by radiologists who know it's a risk factor. It's an "ethi-

cal issue," he says. So Cuzick is urging radiologists and surgeons to share breast density information with patients. To circumvent concerns about fostering anxiety, Cuzick is focusing his energies on physicians at U.K. screening clinics where women with suspicious mammograms are referred. These women, Cuzick reasons, are already worried.

His efforts are meeting with mixed results. "Some [physicians] are prepared to do it, and some are saying they can't do it," he says.

Others can appreciate Cuzick's views, but their own are more nuanced. "There's a whole philosophical concern" about communicating with patients before important questions about density are answered, says Martin Yaffe, a medical biophysicist at Sunnybrook and Women's College Health Sciences Centre in Toronto. Tlsty is torn. She'd want her own density measures, she says, but isn't sure density is ready for widespread clinical use.

Many physicians say they'd be more inclined to use density measures if they appeared in risk assessment models. Mitchell Gail, a biostatistician at the National Cancer Institute, is experimenting with adding density to the so-called Gail model, commonly used in the United States. The Gail model already includes such risk factors as age, family history, and age at first menstrual period. Gail is testing whether adding density makes it more precise. (Models don't always need to include all risk factors, especially those that travel together.) But, says Byrne, who's working with Gail, "Nothing's going to tell you yes or no absolutely" about whether you'll contract the disease.

The risk models also can't answer a critical question: Does lowering density lower breast cancer risk? Malcolm Pike, a cancer

epidemiologist at the University of Southern California, argues that it does, given the drug tamoxifen's ability to lower breast density as well as breast cancer risk—or, for that matter, the tendency of hormone replacement therapy to increase density alongside cancer risk.

But Boyd, Yaffe, and others are not so sure. "Just because two things move together doesn't mean that one thing causes the other," says Yaffe. Testing the theory is difficult, though Boyd is trying with a diet study that will wrap up later this year. There, 4700 women are assigned to normal or low-fat diets for at least eight years. After two years, the low-fat diets reduced density slightly; Boyd doesn't yet know whether the diets also lower the chance of breast cancer.

Given all these uncertainties, perhaps the biggest question for doctors and patients is what women with dense breasts can do about it. Extra screening is one obvious possibility—except that the most common and cheapest screening tool, mammography, isn't always well suited to catch tumors in dense tissue. (MRIs and ultrasounds can, however.) Boyd's evidence that diet can alter density hasn't been replicated. Another controversial option is preventive drugs: Tamoxifen has been shown to stave off breast cancer, and two other drugs, raloxifene and lasofoxifene, are being tested in large trials nearing completion. (Both are designed to treat and prevent osteoporosis.)

But some oncologists are uncomfortable considering preventive drugs for women whose only breast cancer risk factor is high density. Numbers alone make density delicate to use in a chemoprevention setting. If breast density were treated like any other risk factor, that would make 15% to 20% of postmenopausal women eligible for preventive drugs, estimates Cummings, who has worked closely with Eli Lilly and Pfizer, the drugs' makers. It's also not clear whether prevention drugs act along the same pathway as density, says Boyd, and thus how effective they would be in women with high density.

"We need to be very careful in saying, oh, you have 75% breast density," says Carol Fabian, an oncologist who directs the breast cancer prevention program at the University of Kansas Medical Center. "You need to be on tamoxifen."

Cancer link? Like tumors, dense breast tissue (*above*) is packed with collagen, shown here in pink.

–JENNIFER COUZIN

With reporting by former intern Cathy Tran.

Deep Impact Finds a Flying Snowbank of a Comet

Blasting a comet for science's sake turns out to be harder than expected, but targeting Tempel 1 has revealed a fragile body that's already been through geologic turmoil

Before comet researchers smashed a nearly half-ton chunk of metal into Tempel 1 last July, they weren't at all sure how a comet would react to such an assault (*Science*, 27 May, p. 1247). The possibilities ranged from imperceptible to apocalyptic. Perhaps it would swallow the impactor without a trace, like a marshmallow engulfing a BB. Perhaps it would form a fairly conventional impact crater. Or perhaps it would go completely to pieces. Just what kind of hole the impact left in the 14-kilometer-long chunk of ice and dirt should tell what this comet is like, Deep Impact team members said. Such groundbreaking exploration, they added, would no doubt surprise them.

They got the surprise right. The researchers report online in *Science* this week* that the impact blew out so much unexpectedly fine powder that they still haven't seen the crater through all the dust. Indeed, they might well never extract the crater from the images returned by the Deep Impact mother craft.

Even so, analysis of close-up observations of the surface and the way impact ejecta behaved shows that comets "are not

Destination Tempel 1. Smooth areas edged by lighter scarps (*top and lower left*) suggest that the comet has been geologically reworked.
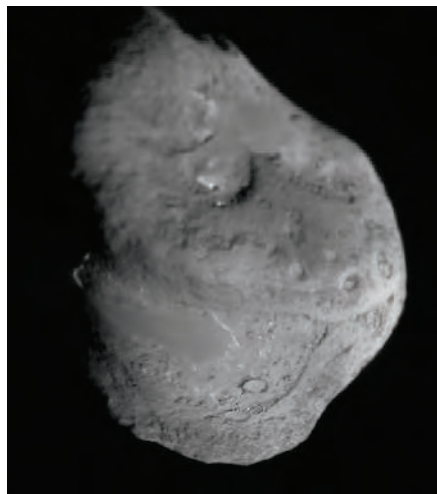
though, with powder-size particles weakly agglomerated into something with the consistency of a snowbank. And, much to everyone's surprise, Tempel 1 shows signs of past geological activity; it is not just a primordial dirty snowball.

Several sorts of observations made the inferences possible. The spacecraft's

weak, fine-grained comet. The spacecraft imaged an expanding cone of ejecta as close to the crater as it could view. The cone continued to rise for more than an hour, which means the outer meters of the comet must consist of weak, easily powdered material. The rate at which the ejecta cone expanded near its base depends on the strength of Tempel 1's gravity. From that dependence the team calculates that the comet as a whole has a density of roughly 0.6 grams per cubic centimeter, two-thirds that of pure water ice. Something like 50% to 70% of Tempel 1 must therefore consist of empty space, say Veverka and mission principal investigator Michael A'Hearn of the University of Maryland, College Park. "The closest [analog] may be a fresh, light New England snowfall, except dirty," says team member Peter Schultz of Brown University in Providence, Rhode Island.

A cruddy flying snowbank isn't that far from what planetary scientists foresaw, but they did expect to find that comets have been flying around pretty much unaltered since they formed 4.5 billion years ago. Close-up looks at comets Halley, Borrelly, and Wild 2 (*Science*, 5 October 2001, p. 27; 9 January 2004, p. 151) have revealed the expected surficial battering by impacts and craggy decay from loss of ice to solar heating. But Tempel 1 also has dramatic examples of layers, geologic strata of uncertain origin. "We're still puzzling through the layering," says A'Hearn. One smooth area appears eaten away at its edges, revealing an older layer beneath with its own muted impact craters. No one is willing to say whether the layers formed when Tempel 1 did, or much later. They might have been laid down as the comet formed layer upon layer 4.6 billion years ago. The layers are reminiscent of those recently found on Saturn's cometlike moon Phoebe (*Science*, 18 June 2004, p. 1727).

"Comet nuclei have clearly undergone geologically interesting processes," says Veverka. That news excites the geologists, but it could frustrate the geochemists looking for unaltered cometary material in the Deep Impact debris. Their search has just begun.

–**Richard A. Kerr**
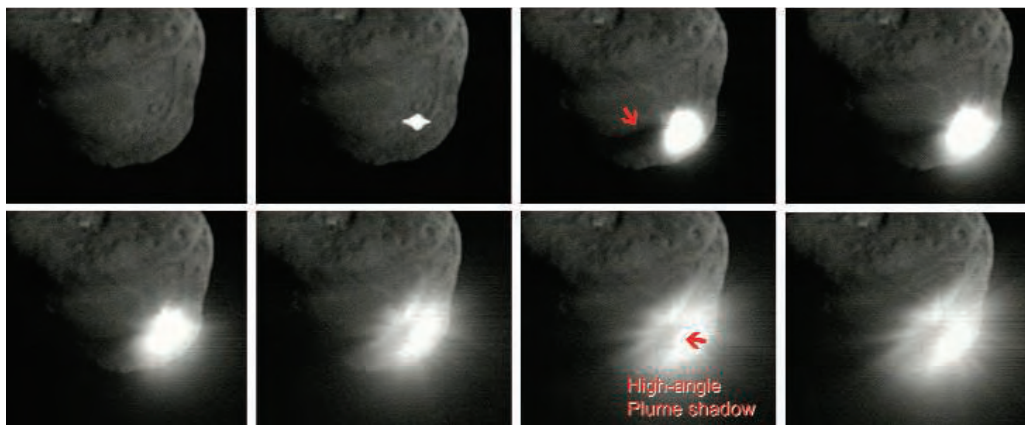
Dustup. The impactor threw up a surprising amount of powdery debris, so much that a rising dust plume cast a shadow (red arrows) more than 5 seconds after impact and obscured the resulting crater.

cosmic candy fluff or chunks of concrete," says team member Joseph Veverka of Cornell University in Ithaca, New York. This one at least is decidedly on the fluffy side,

infrared spectrometer measured varying surface temperatures under different illumination by the sun. The readings showed that the surface is quick to warm or cool, which means it must be porous like loose sand or granular snow, not a solid block of ice.

What the mother ship could see of the impact and its aftermath also points to a

CREDITS: NASA/JPL-CALTECH/UMD

Coming Into Focus: A Universe Shaped By Violent Galaxies

Astronomers and theorists still lack a detailed model of how the Milky Way and its cousins formed. But they now agree on one thing: Galaxies really blow

SANTA CRUZ, CALIFORNIA—When theorists try to simulate how galaxies grow, they never quite get it right. In the computer world, spirals like our Milky Way rarely develop graceful disks with the sizes that astronomers see. The giant blobs called elliptical galaxies spawn new stars for too long. And models predict too many tiny galaxies, strewn through volumes where astronomers see mostly blackness.
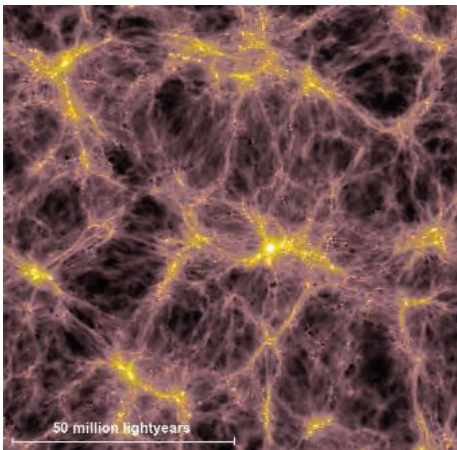
Explosive spread. In a simulation of gas contained in galaxies 2.2 billion years after the big bang (*left*), winds from supernovas propel heavy elements deep into space (*right*).

But in the last year, modelers have improved their galaxy recipes by adding generous portions of wanton violence, which alters the appearance and dynamics of galaxies in just the right ways. At all stages of the universe's growth, blasts from supernovas and surges of energy from massive black holes appear to dictate the fates of galaxies. These waves of unrest—often sparked by galactic collisions—wreak havoc on a galaxy's supply of star-forming gas but in dramatically different ways for big galaxies and small ones.

Astronomers have learned to recognize this "feedback" in galaxies, and they now see it nearly everywhere they look. What's more, simulations that include feedback and a healthy dollop of galaxy mergers are producing the most realistic assemblies of virtual galaxies yet created. "Feedback is really the new frontier in understanding galaxy formation," says astronomer S. Michael Fall of the Space Telescope Science Institute in Baltimore, Maryland. Still, the gaseous interplay

that drives feedback is so complex—and so poorly understood—that no one yet considers the galaxy problem solved.

The dark new framework

Feedback in various guises suffused a crowded agenda for 250 scientists meeting* last month among the redwood trees of the University of California (UC), Santa Cruz. A bracing blend of leading theorists and observers focused on galaxies shining within a hidden network of dark matter that spans the cosmos.
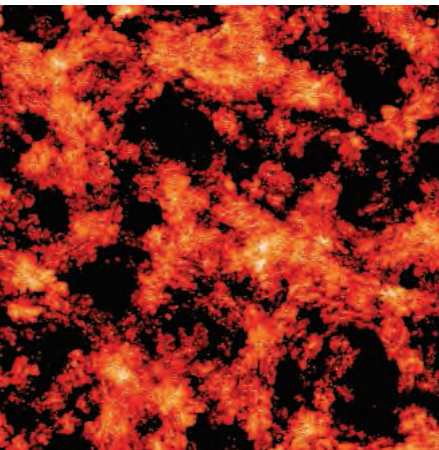
For more than two decades, astronomers have known that clumps of dark matter create the gravitational pits within which galaxies assemble. "It's almost a standard model," says astrophysicist Joseph Silk of the University of Oxford, U.K. Indeed, the meeting celebrated the 60th birthdays of three UC Santa Cruz researchers who first conceived that model in a 1984 paper in *Nature*: astrophysicist George Blumenthal, astronomer Sandra Faber, and cosmologist Joel Primack. (The fourth author, astronomer Martin Rees of the University of Cambridge, U.K., could not attend.)

The universe has grown stranger since then. Cosmologists now believe that "dark energy," an utterly unknown force, drives an ever-accelerating expansion of space

---

* Nearly Normal Galaxies in a Lambda-CDM Universe, 8–12 August 2005

(*Science*, 2 September, p. 1482). But dark energy won little overt notice at the meeting, for it becomes important only on scales far larger than an individual galaxy. "It's the greatest discovery of the century, yet it has had very little effect on people like us," Faber mused.

However, dark energy did set a framework in two areas. First, several speakers examined the history of galactic mergers. Nearby collisions fling arcs of stars into space and spark telltale bursts of starbirth. The signs aren't as clear at great distances, where astronomers try to reconstruct the younger universe from faint images. "Distant galaxies all look a little funky," says astronomer Hans-Walter Rix of the Max Planck Institute for Astronomy in Heidelberg, Germany. With no clean way to spot mergers, he says, astronomers can't specify how much a given galaxy transformed itself by devouring its neighbors. But they suspect the collisions happened often, especially in rich galaxy clusters.

Galactic makeovers are rarer today, says Matthias Steinmetz, director of the Astrophysical Institute Potsdam in Germany. The main culprit is dark energy, which has pushed galaxies farther apart than they otherwise would be. As a result, the universe of the last 7 billion years—roughly half the cosmic age—has provided a comparatively peaceful setting for galaxy growth. Many spirals today might not exist without dark energy, Steinmetz notes, because mergers damage their extended flattened disks.

Still, astronomers have scoured space for imprints of mergers that do continue. New research suggests that elliptical galaxies—the behemoths of the cosmos—are far from inert. Astronomer Pieter van Dokkum of Yale University in New Haven, Connecticut, and colleagues took sensitive images of 126 ellipticals to search for disrupted patterns. Wide fans of stars and other collision remnants popped out for more than half of the galaxies. Van Dokkum's analysis suggests that elliptical galaxies add 10% more mass to their bloated retinues every billion years.

"It's clear that a picture of galaxy formation without mergers as a central theme does not work," says Rix.

The red-and-blue divide

These greedy acquisitions are just one feature of elliptical galaxies. More puzzling is how ellipticals produced dazzling bursts of new stars in the young universe—as early as a billion years after the big bang—and then shut down just as quickly (*Science*, 19 March 2004, p. 1750). Galaxy modelers have now generated results that look tantalizingly close to real observations.

Researchers can also thank dark energy

for this advance—at least indirectly. "It used to be that whenever you did a simulation, you had to run many of them because you didn't know the cosmological parameters," says meeting co-organizer Avishai Dekel of Hebrew University in Jerusalem, Israel. "Now that we know [the proportions of] dark matter and dark energy, most effort goes into the physics of galaxy formation."

The progress is helping astronomers understand why galaxies occupy two broad color categories, like the states in U.S. electoral politics. Elliptical galaxies shine with the cool reddened light of ancient stars, with virtually no gas left to make more suns. On the other hand, disk galaxies brim with gas and sparkle with new stars, blazing with blue and ultraviolet vigor. Red galaxies can grow 10 to 100 times as massive as their blue cousins can. Ongoing surveys of tens of thousands of galaxies also show that red ones typically clump together, while blue ones spread far and wide across space.

An ambitious new simulation has reproduced these galaxy properties, and many others. Astrophysicists led by Volker Springel and Simon White of the Max Planck Institute for Astrophysics in Garching, Germany, created the Millennium Run: a model containing 10 billion particles of dark matter, arrayed within a cube more than 2 billion light-years on a side. The team started with minuscule fluctuations in the distribution of matter, as reflected in the subtle patterns of the cosmic microwave background—the remnant heat of the big bang. Then, gravity and dark energy acted on those fluctuations for the age of the universe in simulated time. The result was a striking web of dark matter, bearing eerie resemblance to a neural network.

Next, the team used a separate model of gas physics, stars, and the dynamics of massive black holes to track simulated galaxies within each knot of dark matter—about 18 million in all. As Springel and White reported at the meeting and in the 2 June issue of *Nature*, the physical properties of these galaxies neatly captured the red-and-blue divide. Notably, the old, massive red galaxies clustered tightly within the densest regions of dark matter, while the youngest blue galaxies spread out smoothly—the first such match for a large-scale model. "To me, this effect was quite unexpected," says White.

The monumental effort won fans, including Primack, who has modeled the growth of cosmic structure since the mid-1980s. "I'm

The rich get richer. Streamers and fans of stars show huge elliptical galaxies have absorbed smaller companions.

absolutely blown away by the success of these huge simulations," he says.

Furious outpourings of energy were the key to making the model work, says Springel. "Feedback is essential for structure formation on all scales," he notes. For instance, shock waves and winds blown by supernovas will churn up small to medium-sized galaxies and prevent them from forming as many stars as they would otherwise. This effect can short-circuit star formation in some dwarf galaxies before it takes hold, leaving nuggets of dark matter with little visible light.

In bigger galaxies, even supernovas don't have enough oomph to propel gas into deep space. Instead, powerful beacons of radio energy—driven by matter cascading into supermassive black holes—rule the outcome. When galaxies reach a critical mass, this feedback heats gas so intensely that new stars can no longer form. The galaxies become "red and dead," massive old ellipticals sitting at the hearts of giant galaxy clusters.

Prepare for blowout

Other new models zeroed in on this "blowout phase." In a simulation led by graduate student Philip Hopkins and astrophysicist Lars Hernquist of the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, galaxy cores turn "on" and "off" more erratically than thought. When two gas-rich galaxies collide, gravitational forces propel breathtaking streams of gas toward the newly merged

Lost in space. Stellar outbursts push hot gas out of a nearby dwarf galaxy (white outline).

black hole at the center. The hot disk of matter around the black hole emits as much radiation as physics allows, creating a hyperluminous quasar.

Movies of this process show that the quasar expels gas from the merging galaxies in what amounts to a detonation. It's aberrant behavior for an object that is relatively quiet for perhaps 99% of its history. Indeed, nearly every galaxy may have hosted a bright quasar throwing at least one such tantrum, Hernquist says. The model successfully explains the numbers of active quasars, how red galaxies evolve, and a strict relationship between the mass of a galaxy's central black hole and the spheroidal bulge of stars around it. The team has submitted its work to the *Astrophysical Journal*.

Studies of galaxies on all scales bolster these models. For instance, x-ray images of nearby dwarfs show that some of them seem turned inside out by supernovas and winds from massive stars, ejecting gas beyond the galaxies' reach. And at the meeting, astronomer Charles Steidel of the California Institute of Technology in Pasadena reported on studies of the environments around extremely distant galaxies with one of the 10-meter Keck Telescopes in Hawaii. Rapid motions of heavy elements near the galaxies point to outward wind speeds of 500 to 600 kilometers per second, with much of the gas escaping the galactic clutches. In extreme cases, Steidel's team found quasars that have swept out space around them for tens of millions of light-years.

Speakers agreed that the models and observations point to many galaxies growing in fits and starts, ultimately limited by violent feedback. But galactic gas is complex, and it's not clear how it responds to winds, shocks, and explosions. For example, most models at the meeting included a prescription that a galaxy of a certain mass will expel a certain percentage of its gas as it forms. "But there's no understanding of why," says physicist Anthony Aguirre of UC, Santa Cruz. "For everything on the scale between a star and a galaxy, we're more or less clueless. It's really the gas that makes things bad."

For this reason, the impressive simulations do not yet offer a convincing explanation of what makes galaxies tick. As Oxford's Joseph Silk cautioned about the Millennium Run: "It is a beautiful picture of how [large-scale] structure forms, but is it right?" The answer, he says, awaits a deeper scrutiny of the physics of gas and stars close to home.

–ROBERT IRION
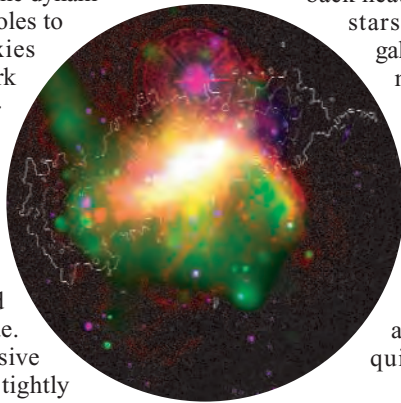
Random Samples

Edited by Constance Holden

Out in The Cold

Florida manatees will soon need a new winter haven. For nearly 50 years, when ambient water temperatures drop below 20°C, manatees have basked in the warmth of outflows from nearly a dozen power plants on the Atlantic and Gulf coasts.

Now, stringent environmental regulations and rising fuel costs may shut many of the plants down. Weaning the animals off the outflows must

Manatees laze in power plant water.

begin soon, say David Laist of the Marine Mammal Commission in Bethesda, Maryland, and John E. Reynolds of the Mote Marine Laboratory in Sarasota, Florida, in the summer issue of *Coastal Management*. If a plant shuts down during a cold winter, hundreds of manatees with no alternate sanctuary could die, they warn. "Once we identify solutions that could work, it will take a long time to get them into place," Reynolds says.

So how to teach an old manatee new tricks? Reopen access to the natural springs that used to shelter them and create solar-power-heated refuges to tide them over until they find their new digs, say the researchers.

Plying an Ancient Trade Route

A 12.5-meter reed boat, a replica of crafts that plied the Persian Gulf 5000 years ago, has been reconstructed by marine archeologists. Composed only of traditional materials like reeds, tar, leather, goat hair, and ropes of date palm fibers, the boat— named *Magan*—was to set sail on 7 September from the port city of Sur in Oman for the Indian port of Bet Dwarka in Gujarat.

The international 8-member crew, fueled by $200,000 from the government, will ride the monsoon winds with the intent of following the ancient ocean trade route thought to have linked the great civilizations of Mesopotamia and the Indus Valley. Historians dispute whether India and Oman actively traded in this era, but

Indian artifacts such as seals, pottery, and beads have been found in Oman.

Marine archaeologist Alok Tripathi of the Archaeological Survey of India, New Delhi, says navigation will be done using only the sun and stars. "We hope to learn how Bronze Age mariners coped with oceans," he says. And the crew will eat what crews of yore ate: dates, cheese, and dried fish.

Surf 'n' Turf

People living on the coast of Wales 12,000 years ago got about a third of their food from the sea, according to an isotope analysis of their bones. That makes them the earliest people known to have intensively used the ocean for food.

Shells and fishbones have been found at much earlier human sites, but the bones indicate that seafood formed only a small part of the diet.

By comparing isotopes of nitrogen and carbon found in the bones of four individuals found at the cliff site known as Kendrick's Cave, Michael Richards, an archaeologist at Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, and colleagues were able to estimate the ratio of terrestrial to marine food in their diets. The data suggest people

ate predators at the top of the marine food chain—most likely seals, Richards says.

The findings, reported in this month's issue of *Journal of Human Evolution*, reveal an early example of people switching from hunting and gathering to more intensive use of certain food sources, says Richards's team. That was a significant step toward the development of a settled, agricultural lifestyle that began in Britain about 5000 years ago.

The study makes "a real contribution to our knowledge of Late Glacial adaptations," says archaeologist Rich Schulting at Queen's University Belfast in Ireland. "The coastline of the time is gone, so finding evidence for if and how marine resources were used is especially challenging."

Oldest Body to Science

Barely four hours after the oldest women in the world died at a Dutch nursing home on 30 August, scientists at a nearby university lab had performed an autopsy, discovered that she died of stomach cancer, preserved key body parts, and started a battery of tests that they hope will shed light on why she lived so long.

Hendrikje van Andel-Schipper, 115, decided some 30 years ago that she would donate her body to science, says anatomist and neuroscientist Gert Holstege of the University of Groningen.

Although her hearing was poor and her vision almost gone, Van Andel-Schipper had

remained sharp. She was an ardent soccer fan, and, says Holstege, "she listened to the radio news every hour until she was 113." Holstege hopes to find that Van Andel-Schipper's brain does not have the brain lesions often seen in the elderly, even those not suffering from dementia. That would help dispel the idea that neurodegeneration is inevitable, he says.

But aging researcher Thomas Kirkwood of the University of Newcastle upon Tyne, U.K., is skeptical. "One case doesn't tell you very much" beyond the obvious, he says: that Van Andel-Schipper had "an exceptionally good body for a woman her age."

DEATHS

An appetite for knowledge. Horace Davenport, the physiologist who discovered what prevents the stomach from digesting itself, died at his home in Ann Arbor, Michigan, on August 29. He was 92.

Davenport discovered the role of carbonic anhydrase in the parietal cells of the stomach, which led to a better understanding of how the gastro-mucosal barrier works. "Many successful therapies for peptic ulcer today are based on the discovery of this mechanism," says Howard Markel, director of the Center for the History of Medicine at the University of Michigan, where Davenport served as a faculty member until 1983.

Davenport also made a significant contribution to the teaching of physiology by authoring three textbooks, one of which—"The ABC of Acid-Base Chemistry"—has sold over 140,000 copies and is now in its sixth edition.

JOBS

No more waiting. The head of the Office of Women's Health at the U.S. Food and Drug Administration (FDA) has quit over the agency's controversial handling of Plan B, the emergency contraceptive.

Susan Wood, a biologist who joined FDA in 2000, resigned less than a week after the agency announced another delay on whether to sell Plan B over the counter. The drug is currently available by prescription only. FDA Commissioner Lester Crawford has requested public comments on whether to limit unrestricted sales to older teenagers and adult women.

But Wood had had enough. "I can no longer serve as staff when scientific and clinical evidence ... has been overruled," said Wood in her e-mail, which was released by reproductive health groups.

FDA put out a statement calling her departure "unfortunate."

Senators Patty Murray (D–WA) and Hillary Clinton (D–NY) called Wood's departure "disturbing" and said it reflects the politicization of the agency.

DATA POINT

Physics upswing. The number of U.S. citizens beginning graduate school in physics rose by 47% from 1998 to 2003, according to a new report from the American Institute of Physics (AIP) (http://www.aip.org/statistics/trends/reports/ed.pdf). That growing domestic interest contributed to an overall jump of 31%. Foreign student enrollment rose by a more modest 16%, reflecting tougher U.S. visa restrictions after the 2001 terrorist attacks.

The report also documents the first uptick in 8 years in physics Ph.D.s produced by U.S. universities, as well as a 25% jump since 1999—a 4-decade low point—in the number of bachelor's degrees in physics awarded by U.S. institutions. AIP's Patrick Mulvey suggests it's a result of a stronger job market.

First-year U.S. and foreign graduate physics students, 1990 fall–2003 fall

Fueling a dream. "A nuclear rocket is required for humans to go to mars," says nuclear engineer Steven Howe, director of the new Center for Space Nuclear Research in Idaho Falls, Idaho. Howe hopes the center, affiliated with the Idaho National Laboratory and due to open next month, will foster university research to "support the work that will be coming out 10 years from now, hopefully when it's needed" for a Mars mission. But he concedes that nuclear propulsion's future role hinges on NASA's long-awaited exploration plan, expected in September. (In February, the administration cancelled a nuclear-propelled probe mission to Jupiter.) Educating a wary public will be another role for the center, he said.

Howe, 52, has published work ranging from antiproton studies to high-speed aerodynamics during years at Los Alamos National Laboratory. He has also dabbled in science fiction: His novel about a moon base, *Honor Bound Honor Born*, came out in 1997.

PIONEERS

Creating connections. Victoria Gray has spent her career collecting the names of creative individuals from all walks of life. Now she's using those contacts to inspire some of the most promising students in the United States.

Last month, Gray brought four Nobelists, three Pulitzer Prize winners, one U.S. senator, and four dozen other high achievers to the Stanford Linear Accelerator Center (SLAC) in Menlo Park, California, for a weekend with 127 high-school students. The goal of the conference, entitled "Adventures of the Mind," was to help students maximize their potential by acquainting them with the life stories of great achievers, says Gray, who trained as a lawyer and

who lives in Washington, D.C. Invitees including Michael Calderbank of Princeton High School in New Jersey rubbed elbows with the likes of (left to right) Nobelist Leon Lederman, author Amy Tan, and physicist Janet Conrad.

Gray solicits nominations and selects the students for the meeting—the first was held in Seattle in 2003—with an eye to those from lower socioeconomic backgrounds. The funding comes from private donors, and the mentors volunteer their time. "They know that through me they can reach the right kids and make a difference," Gray says, "and have fun."

Women Making Strides in Big Pharma

AS FEMALE SCIENTISTS IN LEADERSHIP POSI-tions at a large pharmaceutical company, we disagree with the conclusion that "It's still a man's world at the top of big pharma research" (J. Mervis, Special Section on Drug Discovery, News, 29 July, p. 724). There are numerous examples suggesting that this notion is more than a bit outdated. Of additional concern is the superficial analysis that led to the conclusion that childcare issues hold women back in the field of discovery research.

Certainly, we would all like to see more women at the heads of research organizations, but we have observed significant progress over the past 10 years and believe that the future looks bright. There are now many women leading the science and influencing research directions in pharmaceutical R&D organizations. They and, increasingly, their male colleagues manage the demands of child-rearing while achieving significant career growth in this profession.

Women across professions—law, academic, science, and corporate—and their families continue to successfully manage the challenges presented by raising children and growing careers at the same time. Women with careers in discovery research are no exception. We believe that pharmaceutical companies, large and small, are great places for women to pursue careers as research scientists and regret that *Science* did not look at this area more broadly before drawing the unfortunate conclusion to the contrary.

CATHERINE D. STRADER,[1] SATWANT K. NARULA,[2] JEAN E. LACHOWICZ[3]

[1]Executive Vice President, Drug Discovery Research, [2]Vice President, Inflammation and Infection Discovery Research, [3]Director, Cardiovascular & Metabolic Diseases Discovery Research, Schering-Plough Research Institute, Kenilworth, NJ 07033, USA.

The Importance of Origins?

IN HIS ESSAY "SCIENCE IN THE ARAB WORLD: vision of glories beyond" (3 June, p. 1416), W. Maziak states that "[o]ne [of the knowledge shocks that ignited the Renaissance] was delivered by Ibn-Sina (Avicenna, 980 to 1037), whose Kitab Al-Shifa ("The Book of Healing") introduced medieval Europe to the principles of logic and their use to gain knowledge, and placed science and religion on equal terms as sources of knowledge and understanding of the universe." However,

Avicenna was not an Arab. He was a Persian scientist who spoke the Persian language as his mother tongue and who wrote in both Persian and in Arabic. Maziak also implies that the well-known Persian scientists and philosophers (Al-) Razi's (Rhazes) (*1*, *2*) and (Al-) Khwarizmi (Kharazmi) (*3*) in the 9th and 10th centuries were also Arabs. Although the genuinely sincere attempts by Maziak to offset contemporary ignorance of and/or bias against the important role played by Asian and Middle-Eastern scientists is to be commended, his misrepresentation of the history and science does not help.

KAMYAR KALANTAR-ZADEH AND MOHAMMAD NAVAB

David Geffen School of Medicine at UCLA, Los Angeles, CA 90095, USA.

References
1. L. Richter-Bernburg, *Med. Secoli* **6**, 377 (1994).
2. L. F. Haas, *J. Neurol. Neurosurg. Psychiatry* **54**, 483 (1991).
3. K. C. Ryding, *Ambix* **41**, 121 (1994).

IN HIS ESSAY "SCIENCE IN THE ARAB WORLD: vision of glories beyond" (3 June, p. 1416), W. Maziak describes the challenges facing the Arab world in their quest to realize scientific prosperity. Maziak makes an error: Razi, Ibn-Sina, and Khwarizmi were Iranian scientists—not Arab scientists.

The "Arab-Islamic" label or even "Islamic" label is also misleading because not all of the scientists of the era were even Muslim (*1*). For example, Khwarizmi was also known as Al-Majusi (the Magus), which suggests that he was Zoroastrian.

One could appropriately argue that ethnicity is unimportant and what matters is the contribution of any scientist to the advancement of knowledge for humankind. However, wouldn't any reader have had a similar reaction while reading an article that described Sir Isaac Newton as a Frenchman and Marie Curie as a Spaniard?

Lumping these scientists into the culturally narrow label of "Arab-Islamic" is historically inaccurate. This label does not recognize the rich diversity of Eastern scholars that contributed to science in an era where science was essentially nonexistent in the Western world and was later invaluable to its Renaissance.

VARSHASB BROUMAND,[1] SALOMEH KEYHANI,[2] BERHOOZ BROUMAND[3]

Nephrology Associates of Northern Virginia, 3700 Joseph Siewick Drive, #305, Fairfax, VA 22033, USA.

[2]Mount Sinai School of Medicine, New York, NY 10029, USA. [3]Iran University of Medical Sciences, Tehran 19549, Iran.

Reference
1. B. Broumand, K. Zandinejad, *Am. J. Nephrol.* **16** (no. 6), 550 (1996).

Response

THE LETTER WRITERS DWELL ON THE NATION-ality and religion of scholars of the golden era of Arab-Islamic civilization. I did not indicate or mean to imply that early scholars of the golden era were all Arabs or Muslims. The figure legend referring to Avicenna as an Arab scientist and philosopher was inserted by the editorial staff, and I did not pay attention to this mistake when I revised the final version.

This Essay, if read as a whole, is a critical account of the current status of science in the Arab world. History was used briefly as a background to show that when tolerance and embracement of science prevailed, science moved forward. The fact that some of the great figures of the golden period were not Arabs or Muslims or worked in other languages such as Syriac, Hebrew, or Persian is a clear demonstration of this notion. I wrote that "scholars of every color and creed traveled to Damascus and Baghdad to study and work." The sociopolitical environment of the state, or for that matter the Abbasid Caliphate, provided the required catalyst for people of every background to investigate and excel. It was within that nourishing and tolerant melting pot that ancient knowledge was preserved and new knowledge was produced, a dynamic that proved instrumental to the European Renaissance. This is the context most relevant, in my opinion, to the prospects of science advancement in any society today, and this is the context I tried to reflect.

When we see what blind politics, business, and ideologies have made of our world today, it becomes clear that cross-national and cross-cultural relations between people of science, arts, and reason have become an urgent need to reduce tension, avert conflicts, and lobby for a more humane world. Sacrificing some of our group-pride, no matter along what lines the group is defined, seems to be the price we ought to pay for a more peaceful and prosper-

Scientist and philosopher Ibn-Sina (Avicenna) surrounded by his students.

ous coexistence. Carl Sagan wrote, "Whenever our ethnic or national prejudices are aroused, in times of scarcity, during challenges to national self-esteem or nerve, when we agonize about our diminished cosmic place and purpose, or when fanaticism is bubbling up around us—them, habits of thought familiar from ages past reach for the controls. The candle flame gutters. Its little pool of light trembles. Darkness gathers. The demons begin to stir." [(1), pp. 26–27].

WASIM MAZIAK
Syrian Center for Tobacco Studies, Post Office Box 16542, Aleppo, Syria.

Reference
1. C. Sagan, *The Demon Haunted World* (Ballentine Books, New York, 1996).

Firearms, Violence, and Self-Protection

THE ASSOCIATION THAT J. B. BINGENHEIMER *et al.* have found between exposure to firearm violence and subsequent perpetration of violence ("Firearm violence exposure and serious violent behavior," Reports, 27 May, p. 1323) may well reflect a causal effect of prior victimization, but I believe they have misread what is being caused. They classified a subject as a "perpetrator of serious violence" if she or he reported being theatened or attacked by another or had "been in" a gang fight, but also if the subject had "carried a hidden weapon." Their dependent or outcome variable is problematic partly because it makes no distinction between defensive, even lawful, violence and offensive violence. More importantly, this variable probably reflects just one type of behavior, carrying weapons for self-protection, which should not be described as violent behavior.

The authors report that 12.6% of their sample of (roughly high school age) youth were classified as perpetrators of serious violence, but do not say what share of these were so classified solely because the person had carried weapons for self-protection. But there is strong reason to believe that this share is over half and could approach 100%. A survey of Chicago high school students conducted by the U.S. Centers for Disease Control and Prevention [(1), p. 26] in the same year as the present study, 2001, revealed that 21.2% had carried weapons (6.3% had carried guns) in the previous 30 days. Thus, one would expect that at least 21% of the present study's sample would report defensive weapon carrying alone, easily enough to account for all of the 12.6% classified as "violent perpetrators."

If most of the variation in the outcome variable is really variation in defensive weapon carrying, it means that all the authors have really discovered, or rediscovered, is the rather banal fact that people who have reasons to believe they are likely to be victimized in the future are more likely to carry guns for defensive purposes (2).

GARY D. KLECK
College of Criminology and Criminal Justice, Florida State University, Tallahassee, FL 32306, USA.

References
1. Centers for Disease Control and Prevention, "Youth risk behavior surveillance—United States, 2001," *Morbid. Mortal. Weekly* **51**, 1 (28 June 2002).
2. G. Kleck, M. Gertz, *J. Res. Crime Delinquency* **35** (no. 2), 193 (1998).

Response

WE APPRECIATE KLECK'S THOUGHTFUL LETTER. It is true that carrying a concealed weapon was the most commonly reported of the behaviors that make up our measure of violent behavior. Nearly 10% of adolescents in our sample reported carrying a concealed weapon in the year prior to their Assessment 3 interview, compared with 4% who reported participating in a gang fight, 2% who reported attacking someone with a weapon, 1% who reported shooting at someone, and less than 1% who reported shooting someone.

Contrary to Kleck's conjecture, however, we believe that it is reasonable to include carrying a concealed weapon in our index of serious violent behavior. Weapon carrying is a logical prerequisite to several explicitly violent acts and is indicative of a certain degree of willingness or intent to engage in violence. Moreover, carrying a hidden weapon is strongly associated with all of the other behaviors included in our measure. Compared with those who denied carrying a hidden weapon, subjects who reported doing so were over 23 times as likely to report attacking someone with a weapon, some 27 times as likely to report shooting at someone, and nearly 10 times as likely to report being in a gang fight. Although almost 70% of those who reported carrying a concealed weapon also reported another violent behavior, less than 3% of those who denied carrying a concealed weapon reported other forms of violence.

Furthermore, the relationships we found between exposure to firearm violence and our index of violent behavior are also evident when

each behavior is analyzed as a separate outcome. We reported in our paper that subjects who reported exposure to firearm violence at Assessment 2 were over three times as likely to report some form of violent behavior at Assessment 3 [odds ratio (OR) = 3.71, χ^2 = 41.99, $P < 0.001$]. Breaking this down by behavior, exposed subjects were more likely than unexposed subjects to report carrying a concealed weapon (OR = 3.47, χ^2 = 31.36, $P < 0.001$), being in a gang fight (OR = 3.74, χ^2 = 16.73, $P < 0.001$), attacking someone with a weapon (OR = 9.77, χ^2 = 21.62, $P < 0.001$), or shooting at someone (OR = 11.40, χ^2 = 13.18, $P < 0.001$). Within the analytic propensity strata, those who were exposed to firearm violence were approximately twice as likely to report some form of violent behavior (OR = 2.43, χ^2 = 11.76, $P = 0.001$) and were also more likely to report carrying a concealed weapon (OR = 2.34, χ^2 = 9.00, $P = 0.003$), shooting at someone (OR = 2.78, χ^2 = 1.62, $P = 0.204$), being in a gang fight (OR = 3.27, χ^2 = 7.95, $P = 0.005$), and attacking someone with a weapon (OR = 8.10, χ^2 = 9.06, $P = 0.003$). These unadjusted and propensity-stratified analyses show that our original results apply not only to weapon carrying but also to other forms of violent behavior.

As Kleck suggests, some of the violent behaviors reported by participants in our study may have been motivated in part by self-defense. Yet motivations can be complex, and people often delude themselves and others about the reasons for their actions. Ours is a study of behavior, and our data on adolescents living in Chicago in the late 1990s strongly support the conclusion that exposure to firearm violence increases the likelihood of violent behavior.

JEFFREY B. BINGENHEIMER,[1] ROBERT T. BRENNAN,[2] FELTON J. EARLS[2]

[1]Department of Health Behavior and Health Education, 1420 Washington Heights, University of Michigan School of Public Health, Ann Arbor, MI 48109–2029, USA. [2]Department of Social Medicine, Harvard Medical School, 1430 Massachusetts Avenue, 4th Floor, Cambridge, MA 02138, USA.

The Sight of Violence and Violent Action

IN THEIR REPORT "FIREARM VIOLENCE EXPOsure and serious violent behavior," (Reports, 27 May, p. 1323), J. B. Bingenheimer *et al.* show a causal effect between exposure to firearm violence and subsequent perpetration of violence. This result was foreseen by Shakespeare nearly 400 years ago: "How oft the sight of means to do ill deeds/Makes ill deeds done!" (*1*). Almost three millennia ago, Homer, too, reached a similar conclusion:

"[T]he very presence of a weapon provokes a man to use it" (*2*). It appears as though the mind files away any "useful" information (device or action), and recalls its "usefulness" at appropriate times.

NARENDRA G. MEHTA

3B/33 Takshila, Off Mahakali Caves Road, Andheri (east), Mumbai 400 093, India. E-mail: ngmehta@rediffmail.com

References
1. W. Shakespeare, *King John*, Act IV, Scene II.
2. Homer, *The Odyssey*, translated by E. V. Rieu (Penguin, New York, 1992), p. 252.

Mistakes in a Map

THE ESSAY "ASCENT OF NANOSCIENCE IN China" by C. Bai (1 July, p. 61) is an articulate overview of the main achievements and progress of nanoscience and nanotechnology in China. On page 61, there is a map with a patch highlighted in orange. This is apparently meant to be a map of China, but there are several errors. Mongolia is highlighted, but it has been an independent country since 1921. Hainan and Taiwan are not highlighted on the map, but they are both provinces of China. I am quite surprised at finding these errors in *Science*.

LEI FU

Institute of Chemistry, Chinese Academy of Sciences, Beijing 100080, China. E-mail: fulei@iccas.ac.cn

Response

THE EDITORS APOLOGIZE FOR INCLUDING A map in which Mongolia is highlighted in such a way as to indicate that it is a part of China. As for the non-highlighting of the islands of Taiwan and Hainan, we assure readers that this does not indicate a *Science* policy on the Taiwan question. *Science* has no such policy.

DONALD KENNEDY

....

TECHNICAL COMMENT ABSTRACTS

COMMENT ON "Grain Boundary Decohesion by Impurity Segregation in a Nickel-Sulfur System"

W. T. Geng, J.-S. Wang, G. B. Olson

Analysis of the binding energies calculated by Yamaguchi *et al.* (Reports, 21 Jan. 2005, p. 393) shows that their proposed microstructure of sulfur aggregation at nickel grain boundaries is unrealistic. Our analysis shows that a different configuration of segregated sulfur atoms is more stable and that the grain boundary volume expansion is only half of what was originally proposed.

Full text at
www.sciencemag.org/cgi/content/full/309/5741/1677c

RESPONSE TO COMMENT ON "Grain Boundary Decohesion by Impurity Segregation in a Nickel-Sulfur System"

M. Yamaguchi, M. Shiga, H. Kaburaki

We estimated the segregation concentration of sulfur atoms at a nickel grain boundary using the average binding energy of sulfur atoms. Geng *et al.* question our interpretations of the binding energies and suggest that a different configuration of sulfur atoms from the one we proposed is more stable. We show that the two configurations have the same structure and energy.

Full text at
www.sciencemag.org/cgi/content/full/309/5741/1677d

....

CORRECTIONS AND CLARIFICATIONS

News Focus: "Ready or not? Human ES cells head toward the clinic" by G. Vogel (10 June, p. 1534). The story should have stated that work by Fred Gage and Ajit Varki indicates that human ES cells can acquire non-human sugar molecules from "serum replacement" media as well as mouse feeder cells. In that team's work, levels of the foreign molecule decreased, but were not eliminated, by growing the cells with heat-inactivated human serum.

Reports: "The optical resonances in carbon nanotubes arise from excitons" by F. Wang *et al.* (6 May, p. 838). In the sixth line of the abstract, the word "bond" should instead be "band."

TECHNICAL COMMENT

Comment on "Grain Boundary Decohesion by Impurity Segregation in a Nickel-Sulfur System"

Yamaguchi et al. (1) examined the embrittlement of nickel (Ni) by progressively adding sulfur (S) atoms to a grain boundary (GB). From first-principles calculations, they concluded that S atoms tend to aggregate at the GB and that the repulsive S-S interactions induce boundary expansion, thus weakening Ni-Ni binding across the boundary. The agreement between their calculated critical S concentration and the measured data (2) suggests that the GB embrittlement of Ni is due to the aggregation of S segregants. Although we believe that the first-order calculations of Yamaguchi et al. (1) are reliable, we question the interpretations of the calculated binding energies and argue that the distribution of S near the GB remains uncertain.

According to the Yamaguchi et al. calculations, the average binding energy for a S atom on site GB0/GB2 (Fig. 1) with a 100% occupation is $-4.75/-4.66$ eV, and the binding energy drops to -4.23 eV when both GB0 and GB2 sites are fully occupied. With the calculated segregation energy ΔE_{seg}, which is defined as the energy lowering when an impurity moves from inner bulk (binding energy = -2.96 eV/S) to the GB region, Yamaguchi et al. estimated the occupation number using McLean's equation of equilibrium segregation (3), $C_{\text{GB}} = [C_{\text{bulk}}\exp(-\Delta E_{\text{seg}}/RT)] / [1 + C_{\text{bulk}}\exp(-\Delta E_{\text{seg}}/RT)]$, with the impurity concentration in the bulk and the temperature as parameters. They noted that a binding energy of -4.23 eV is large enough for S atoms to segregate fully to the GB0 and GB2 sites, and went on to discuss the volume expansion effect of such GB0-GB2 S combinations.

This conclusion is only valid if the segregation process starts and terminates instantly, and we know that segregation can take hours or days (2). As shown in (1), when GB0 (GB2) sites are occupied by S, the binding energy of GB2 (GB0) sites reduces greatly as a result of the repelling interaction between S atoms. For instance, if an S monolayer is formed at GB0 sites first, then the binding energy for a 1/4 monolayer of S at GB2 sites decreases from -4.67 to -3.45 eV. The occupation probability is only on the order of 1% under the experimental conditions in (2) (T = 918 K and C_{bulk} = 25 atomic parts per million), according to McLean's equation. This means that a high concentration of GB0-GB2 pairs is unlikely to appear at the Ni GB.

To conduct a more comprehensive search for S-S pairs at the NiΣ5 (012) tilt GB, we calculated the binding energy of S in the form of both GB0-GBn (n = 1, 3, 4, 5, 6) and GB2-GBm (m = 1, 3, 4, 5, 6, -2, -3, -4, -5, -6) pairs at one monolayer concentration using the same code [Vienna Ab initio Simulation Package (4)] and parameters reported in (1). Our calculations demonstrate that although the GBn (n = 3, 4, 5, 6) site is still less stable than the GB0 site, the binding energy difference is greatly reduced when GB0 is occupied by S. For example, a GB0-GB3 pair is 0.05 eV less stable than a GB0-GB2 pair; in an isolated 1/4 monolayer, S at the GB3 site is 1.16 eV less stable than at the GB2 site (1). Interestingly, we find that although the binding energy of the GB1 site in an isolated monolayer is only 3.25 eV, it increases considerably to 4.29 eV when all GB2 sites are occupied by S. The binding energy of other sites, on the other hand, increases to lesser extents, and the occupation probability under the experimental conditions remains low. Moreover, we calculated the binding energy of a 1/4 monolayer GB1 progressively added, to a GB2 4/4 monolayer. The binding energies are 4.15, 4.23, 4.26, and 4.55 in sequence. Our first-principles calculations thus suggest the formation of GB2-GB1 S combinations at a high concentration.

Using the same technique as in (1), we evaluated the tensile strength of six cases of S segregation, namely, (i) clean GB; (ii) GB2 4/4; (iii) GB2 4/4, GB1 1/4; (iv) GB2 4/4, GB1 2/4; (v) GB2 4/4, GB1 3/4; and (vi) GB2 4/4, GB1 4/4. The tensile strengths are 26, 16, 14, 11, 7.2, and 3.9 GPa, respectively. The calculated tensile strength for the clean GB (26 Gpa) is the same as that reported by Yamaguchi et al. The decrease in tensile strength is proportional to the increase of the GB2-GB1 S-S pair concentration, and in the range of S occupations from (GB2 4/4, GB1 2/4) to (GB2 4/4, GB1 4/4), strong GB decohesion occurs. The GB displacement (with respect to the clean GB) caused by (GB2 4/4, GB1 4/4) is about 0.6 Å, much smaller than that caused by (GB2 4/4, GB0 4/4) (1.2 Å), which has been shown here to be unstable. Further detailed analysis will clarify whether GB expansion or directional change of chemical bonding is the key to the strong decohesion caused by S aggregation.

Fig. 1. Schematic model of a NiΣ5 (012) tilt GB. The gray and yellow circles represent Ni and S atoms, respectively. The site numbering follows Yamaguchi et al. [see figure 1 in (1)]. a = <02$\bar{1}$>; b = <100>; c = <012>.

W. T. Geng
Department of Physics
Qingdao University
Qingdao 266071, China
E-mail: geng@qdu.edu.cn

J.-S. Wang
QuesTek Innovations
Evanston, IL 60208, USA

G. B. Olson
Department of Materials Science and Engineering
Northwestern University
Evanston, IL 60208, USA

References
1. M. Yamaguchi, M. Shiga, H. Kaburaki, *Science* **307**, 393 (2005).
2. J. K. Heuer, P. R. Okamoto, N. Q. Lam, J. F. Stubbins, *J. Nucl. Mater.* **301**, 129 (2002).
3. D. McLean, *Grain Boundaries in Metals* (Oxford Univ. Press, London, 1957).
4. G. Kresse, J. Furthmuller, *Phys. Rev. B* **54**, 11169 (1996).

Response to Comment on "Grain Boundary Decohesion by Impurity Segregation in a Nickel-Sulfur System"

We recently concluded that the strong decohesion of a nickel (Ni) grain boundary (GB) is caused by the aggregation of sulfur (S) atoms on the GB, which repel each other (*1*). We further estimated the segregation concentration using the average binding energy of S atoms. However, Geng *et al.* (*2*) claim that the binding energy should be calculated not on average but incrementally (or sequentially). As they point out, the incremental binding energy ($-E_b^{incr}$) when the GB2 1/4 monolayer is added to the GB0 4/4 monolayer is 3.45 eV/S, as shown in Table 1(I). The occupation possibility of this arrangement is less than 1%, according to the McLean's curve, at 918 K and 25 atomic parts per million, as shown in figure 2A in (*1*).

However, if we take another path of increasing occupation up to the (GB0 4/4, GB2 4/4) configuration as shown in Table 1(II), the minimum binding energy becomes 3.79 eV/S, the occupation of which is about 50%. Geng *et al.* miss the point that there are many paths to the (GB0 4/4, GB2 4/4) configuration. Only a few of all possible paths have been investigated at this stage.

Some experiments indicate that S atoms can segregate up to about two monolayers (14 atomic % S within 5 Å from the GB plane, 14.4 atoms/nm²) (*3–6*). The two-monolayer (GB0 4/4, GB2 4/4) segregation can occur in our consideration using the average binding energy. However, neither the average binding energy nor the incremental binding energy has a theoretical basis, because the binding energy in McLean's equation does not take into account any atomic interactions. For quantitative estimation, the interactions among S atoms for all possible configurations should be taken into account in a statistical method like Monte Carlo simulation (e.g., Metropolis method).

Geng *et al.* (*2*) performed calculations using the same code and parameters as in (*1*). They claim that another two-monolayer segregation (GB1 4/4, GB2 4/4) that has a small GB expansion (0.6 Å) is much more stable than the (GB0 4/4, GB2 4/4) segregation, which has a large expansion (1.2 Å). They suggest that the directional change of chemical bonding in the (GB1 4/4, GB2 4/4) case, rather than the GB expansion in the (GB0 4/4, GB2 4/4) case, causes the strong decohesion. However, we have reexamined such calculations and found that the two configurations have the same structure and energy.

Table 1(III) shows our calculations as well as those of Geng *et al.* and reveals a large discrepancy in the incremental binding energy ($-E_b^{incr}$). For example, the energies for the first added GB1 1/4 monolayer to the GB2 4/4 monolayer are 4.15 eV/S in their calculation and 3.49 eV/S in our calculation. Our results indicate that the stability of the (GB1 4/4, GB2 4/4) case is almost the same as the (GB0 4/4, GB2 4/4) case, as shown in Table 1 (I and III). The average binding energies ($-E_b^{ave}$) for both cases are also almost the same. This seems reasonable, because the relaxed atomic geometries near the GB are almost the same between the two configurations (Fig. 1). In contrast, Geng *et al.* show neither the relaxed atomic structure nor the electron density distribution.

Geng *et al.* claim that the GB expansion of the (GB0 4/4, GB2 4/4) case (1.2 Å) is twice as large as that of the (GB1 4/4, GB2 4/4) case (0.6 Å). Here, they may miscalculate the GB expansion as well as the binding energies. The S atoms enter the GB0 interstitial vacancies in the (GB0 4/4, GB2 4/4) case, whereas S substitutes for Ni at GB1 sites in the (GB1 4/4, GB2 4/4) case. The formula used to cal-

Table 1. The calculated binding energies per one S atom. Two kinds of binding energy are shown: the average binding energies ($-E_b^{ave}$) for all the segregated S atoms, and the incremental binding energies ($-E_b^{incr}$) when S atoms are added one by one. The four equivalent sites are distinguished by the coordinate in the <100> direction. (a, 0.0; b, 0.25; c, 0.50; d, 0.75). See figure 1B in (*1*).

Average	This work $-E_b^{ave}$ (eV/S)	Geng *et al.* $-E_b^{ave}$ (eV/S)
GB0 2/4 (a,c)	4.74	
GB0 4/4	4.75	
GB1 4/4	3.33	3.25
GB2 4/4	4.66	
GB3 4/4	3.63	
GB0 4/4 GB2 4/4	4.23	
GB1 4/4 GB2 4/4	4.24	
GB1 4/4 GB3 4/4	4.10	

Increment	$-E_b^{incr}$ (eV/S)	$-E_b^{incr}$ (eV/S)
(I) (GB0 4/4) → (GB0 4/4, GB2 4/4)		
+ GB2a	3.45	
+ GB2b	3.53	
+ GB2c	3.70	
+ GB2d	4.16	
(II) (GB0a, GB0c) → (GB0 4/4, GB2 4/4)		
+ GB2a	4.17	
+ GB2c	4.07	
+ GB2b	3.86	
+ GB0d	3.79	
+ GB0b	4.35	
+ GB2d	4.15	
(III) (GB2 4/4) → (GB1 4/4, GB2 4/4)		
+ GB1a	3.49	4.15*
+ GB1b	3.42	4.23*
+ GB1c	4.09	4.26*
+ GB1d	4.17	4.55*
Ave.(a–d)	3.79	4.29
(IV) (GB3 4/4) → (GB1 4/4, GB3 4/4)		
+ GB1a	4.33	
+ GB1c	4.23	
+ GB1b	4.87	
+ GB1d	4.83	
Ave.(a–d)	4.57	

*Geng *et al.* do not show the site positions in their Comment.

TECHNICAL COMMENT

Fig. 1. The relaxed atomic geometries by force minimization for the left (GB0 4/4, GB2 4/4) and the right (GB1 4/4, GB2 4/4) cases. The gray and yellow circles represent Ni and S atoms, respectively. We can see that the two geometries near the GB are almost the same. The left case includes 21 layers (Ni, 19; S, 2); the right includes 20 layers (Ni, 18; S, 2). For the left and right cases, the correspondences are as follows: GB0(S)↔GB2(S), GB2(S)↔GB1(S), GB1(Ni)↔GB3(Ni), GB3(Ni)↔GB-2(Ni), and so on.

(GB0 4/4, GB2 4/4)   (GB1 4/4, GB2 4/4)
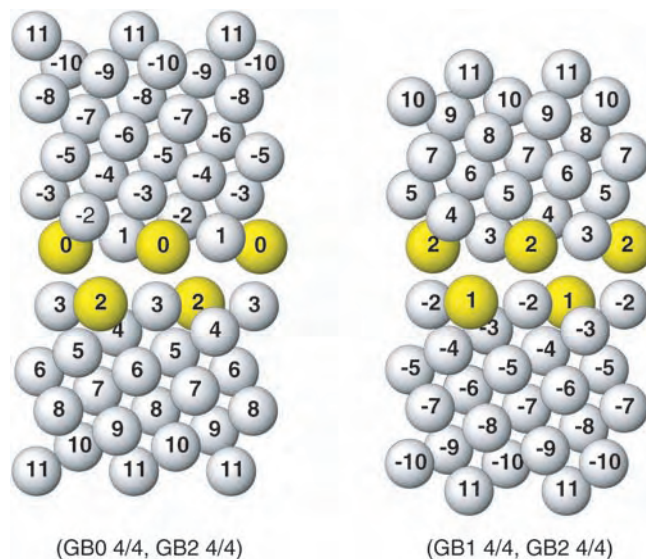
culate the GB expansion [Supporting Online Material in (*1*)] cannot be used to compare the two cases that have different numbers of Ni layers; the discrepancy of the GB expansion (0.6 Å) in the Geng *et al.* calculations may come from the difference in the number of Ni layers, as shown in Fig. 1. The Ni interlayer distance in the <012> direction is 0.78 Å, which agrees well with the difference.

As pointed out by Geng *et al.*, the GB3 monolayer is greatly stabilized by adding the GB1 monolayer. As shown in Table 1(IV), the incremental binding energy of adding GB1 1/4 monolayer to the GB3 4/4 monolayer is about 4.2 to 4.9 eV/S, which is much larger than the average for the GB3 4/4 monolayer (3.63 eV/S) and the GB1 4/4 monolayer (3.33 eV/S). Although the occupation possibility of the GB1 4/4 or the GB3 4/4 monolayer is only

about 1%, all possible paths to the (GB1 4/4, GB3 4/4) segregation should be investigated in detail. Even if this segregation is realized, the final tensile strength is about 12 GPa, which is much larger than the (GB0 4/4, GB2 4/4) and (GB1 4/4, GB2 4/4) cases (<5 GPa).

Contrary to the claim of Geng *et al.*, we cannot find any reason why the (GB1 4/4, GB2 4/4) configuration is much more stable than the (GB0 4/4, GB2 4/4) configuration. Further details are required to make a convincing argument.

*Masatake Yamaguchi**
Motoyuki Shiga
Hideo Kaburaki
Center for Promotion of Computational Science and Engineering
Japan Atomic Energy Research Institute Tokai-mura, Ibaraki-ken, 319-1195, Japan
**E-mail: yamagu@popsvr.tokai.jaeri.go.jp*

References and Notes
1. M. Yamaguchi, M. Shiga, H. Kaburaki, *Science* **307**, 393 (2005).
2. W. T. Geng, J.-S. Wang, G. B. Olson, *Science* **309**, 1677 (2005); www.sciencemag.org/cgi/content/full/309/5741/1677c.
3. J. K. Heuer, P. R. Okamoto, N. Q. Lam, J. F. Stubbins, *J. Nucl. Mater.* **301**, 129 (2002).
4. R. A. Mulford, *Metall. Trans.* **14A**, 865 (1983).
5. A. Larere, M. Guttmann, P. Dumoulin, C. Roques-Carmes, *Acta Metall.* **30**, 685 (1982).
6. C. Loier, J. Y. Boos, *Metall. Trans.* **12A**, 1223 (1981).
7. We thank Y. Nishiyama at JAERI for helpful discussions.

Earning from Risks

Donald MacKenzie

Few aspects of financial markets spark more controversy than derivatives: contracts or securities the value of which depends on the price of another "underlying" asset or on the level of an index or interest rate. Used wisely, derivatives transfer risk to those best placed to bear it, but they are also frequently accused of facilitating insidious accumulation of risk. In 1970, markets in financial derivatives were ad hoc and generally low in volume, and to trade many of today's derivatives would have been illegal. At the end of 2004, according to the Bank for International Settlements, the total notional amount of derivatives contracts outstanding worldwide was almost $295 trillion—around $46,000 for every person on Earth.

Fischer Black was an unlikely source of this explosion of global derivatives markets: brilliant but quirky, reserved, nonconformist, and always motivated far more by ideas than by money. Black had a checkered student life, which culminated in a Ph.D. in artificial intelligence and included being jailed briefly after a riot in 1961—a riot by students protesting Harvard's decision to issue degree certificates in English rather than Latin.

In 1965, Black joined the operations research section of the industrial and management consultancy Arthur D. Little, and there he met Jack Treynor. Although always on the fringes of the academic study of finance, Treynor had developed one of its central inspirations: the capital asset pricing model (CAPM). (Nobel laureate William Sharpe also developed the CAPM, independently and more famously.) CAPM is the "revolutionary idea" in the title of Perry Mehrling's fine biography of Black. The model's central intuition is strikingly straightforward. Markets don't reward investors for shouldering the idiosyncratic

risks of the securities issued by a particular company, because those risks can be minimized to almost any desired extent by diversifying one's investments. Instead, markets reward the kind of risk that can't be removed by diversification: the risk of overall market fluctuations. According to the CAPM, the expected return on an asset is linked in a simple, straight-line fashion to the asset's "beta," its sensitivity to these fluctuations.

Mehrling (an economist at Barnard College, Columbia University) skillfully interweaves the stories of Black's career, his sometimes-difficult personal life, and his contributions to finance and to economics—contributions that were imaginative, distinctive, and at times hard to get past journal referees and editors. Black's lodestar was the CAPM. The work to which his name has become attached permanently,

Risky business. Traders in the pits of the Chicago Board of Options Exchange reacted frantically to the stock market crash of 19 October 1987.

the Black-Scholes-Merton model of option pricing, was for him almost a sideline. But it was a sideline that changed the world.

Options are derivatives that give their holders the right, if they choose, to buy (or in an alternative form of the contract, to sell) an underlying asset such as a block of stock at a set price on, or before, a specified expiration date. Characteristically, Black approached the problem by applying CAPM directly. He was then joined by the economists Myron Scholes (who applied CAPM in a different way) and Robert C. Merton (who distrusted the Sharpe-Treynor CAPM and brought to bear the modern, rigorous mathematics of stochastic processes in continuous time). The work won Scholes and Merton the 1997

Nobel Prize in Economics. Despite Black's continuing status as something of an outsider (in 1984, he left a professorship at MIT for the investment bank Goldman Sachs), he would surely have shared the prize had he not died prematurely in 1995.

Black, Scholes, and Merton assumed that the price of the underlying asset follows a lognormal random walk in continuous time (one in which the logarithms of price changes follow the normal distribution), and that the asset and option trade in frictionless markets, in which, for example, buying and selling incurs no transaction costs. The famous product of these assumptions is the Black-Scholes equation,

$$\frac{\partial w}{\partial t} = rw - rx \frac{\partial w}{\partial x} - \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 w}{\partial x^2}$$

in which w is the value of the option, t is time, x is the price of the stock (which is assumed to pay no dividends), r is the riskless rate of interest, and σ is the volatility of the stock price.

The work changed the financial world in at least two senses. First, when Black, Scholes, and Merton published in 1973, their model's fit to real-world patterns of prices was only rough. The fit improved considerably during the 1970s, in part because the model was used in arbitrage (trading that exploits price discrepancies, and in so doing reduces or eliminates them). Black was directly involved in this use. He set up a service selling sheets of theoretical option values to traders, accompanying the sheets with instructions on how to use them to exploit discrepancies.

Second, the Black-Scholes-Merton model altered how market regulators thought about derivatives, helping to reduce the barriers to derivatives trading. At the start of the 1970s, derivatives were often considered to be little more than wagers on price movements. The model—and the torrent of further work it provoked—provided legitimacy by connecting derivatives to rational pricing, hedging, and market efficiency.

The alignment of reality to the Black-Scholes-Merton model was not permanent. After the 1987 stock market crash, the fit between the model and patterns of option prices deteriorated sharply, with the emergence of a "volatility skew" (or "smile") in those prices that persists to this day. The causes of the skew are still poorly understood, but the crash certainly cast severe doubt on the lognormality of price changes.

The reviewer is at the School of Social and Political Studies, University of Edinburgh, Edinburgh EH8 9LL, UK. E-mail: d.mackenzie@ed.ac.uk

CREDIT: JOHN T. GRALAK/© BETTMANN/CORBIS

Indeed, there have been suspicions that the crash was exacerbated by the large-scale adoption of portfolio insurance, an application of option theory. The technique seeks to create a floor to fluctuations in the value of an investment portfolio and requires the sale of stock or of futures as prices fall.

The ending of the fit between model and reality has not ended the model's effects. The methodology underpinning the model (especially in Merton's version) remains central to how market participants price derivatives and hedge the risks those involve; it's hard to imagine the trading of $295 trillion of derivatives without it. The model is still reflected in how sophisticated traders think and talk—for example, in their constant invocation of "implied volatility" (the volatility of an underlying asset consistent with the price of options on it), a notion that did not exist prior to the work of Fischer Black. In telling the story of Black's life, Mehrling casts light on a crucial, underappreciated aspect of the making of our current world.

MEDICINE

Factoring Risks to the Heart

Amalia M. Issa

While posing for a watercolor portrait on 12 April 1945, Franklin Delano Roosevelt collapsed from a massive cerebral hemorrhage. A young cardiologist, Howard Bruenn, was called to his side, but could do little more than watch; the president died within a few hours. Although Roosevelt's personal physician would tell the press that the president's health had been excellent, Bruenn was among those who knew otherwise—his patient's blood pressure had reached 260/150 earlier that year. Roosevelt had all the risk factors we now know to be associated with cardiovascular disease, including hypertension, smoking, and a high-fat diet.

At the time of Roosevelt's death, we knew little about the causes of heart disease and stroke, and there was nothing to be done about escalating blood pressure.

The reviewer is in the Division of Health, RAND Corporation, Santa Monica, CA 90401 and at the UCLA School of Public Health, Post Office Box 951772, Los Angeles, CA 90095–1772, USA. E-mail: aissa1@ucla.edu

Our understanding has greatly improved since then, progress that owes much to the Framingham Heart Study. When the U.S. Public Health Service launched this landmark epidemiological study in 1948, the small Massachusetts town was chosen for its "typical" demographics, its proximity to Boston cardiologists, and the positive reassurances from the local medical society. The pioneer researchers and their successors have now followed the health and lifestyles of the town's inhabitants for nearly 60 years. About every two years, over 5000 Framingham residents have given researchers detailed medical histories, provided blood samples, and subjected themselves to physical examinations. The original study (initially envisioned as a 20-year investigation) has been followed by the Framingham Offspring Study, and, starting in 2002, the Third Generation Study.

In *A Change of Heart*, Daniel Levy (the study's current director) and Susan Brink (a journalist with *U.S. News and World Report*) offer the lay public a journey through the six-decade history of the Framingham study. While data from Framingham have resulted in over 1200 peer-reviewed scientific publications, the book offers stories of the study and the researchers, physicians, community leaders, volunteers, and subjects who have kept it going.

Scattered through the book are important lessons for the prevention of heart disease that have been drawn from the study. These are interspersed among accounts of how the study was conceived and initiated, anecdotes that reveal quirks of the study's researchers and supporters, and personal portraits of some of the early participants. Levy and Brink offer the reader glimpses into the lives of Framingham residents, who willingly served as subjects, and staff, who staked their professional careers on the study. The authors tell how Evelyn Langley, a housewife and volunteer at her children's school, joined the study's Executive Committee of Residents and went door to door to persuade her neighbors to trust the researchers. They describe how directors Roy Dawber and Bill Kannel started with annual examinations and blood tests, added new technologies such as electrocardiograms, and acquired the necessary understanding of epidemiological methods and data handling. They also recount the National Institutes of Health's decision to shut down the study in 1968 (at the point participants

> **A Change of Heart**
> How the Framingham Heart Study Helped Unravel the Mysteries of Cardiovascular Disease
> *by Daniel Levy and Susan Brink*
>
> Knopf, New York, 2005. 268 pp. $26.95, C$37.95. ISBN 0-375-41275-1.

Measuring a risk factor.

were reaching ages when they would face rising rates of death and disease) and how the study was saved by public outcry and a crucial letter to President Richard Nixon from renowned cardiologist Paul Dudley White.

Current interest in "translational medicine," the transformation of experimental results at the laboratory bench into diagnosis and treatment at the patient's bedside, makes the book's timing fortuitous. Epidemiological studies can serve as an important component of translational research, and the importance of well-designed longitudinal studies that follow population-based cohorts cannot be underestimated. Framingham demonstrates that such studies can provide important information on etiology as well as the natural history and course of disease. It also shows that observational studies can generate hypotheses that are testable in clinical trials.

The authors' account of the beginnings of the Framingham study offers valuable lessons for researchers who wish to embark on community-based research. As the study's first director, Dawber had the foresight to obtain participation from the local physicians and to cultivate the community's sense of "ownership" of the efforts. It is clear that, throughout the years, the townspeople formed a special bond with the researchers and the study. Participants played a crucial role in determining the fate of Framingham Genomic Medicine (a for-profit collaboration of Boston University and venture capitalists), which intended to sell study data and DNA samples to pharmaceutical companies. The participants felt the project would betray the altruistic motives of the study, and their strong opposition killed the plan.

The Framingham Heart Study contributed considerable evidence about the important risk factors for cardiovascular disease, effective lifestyle adjustments, and preventative steps. Well written and backed by citations to primary sources, the engaging account Levy and Brink provide in *A Change of Heart* is appropriate for both professionals and the interested public.

Pathogen Surveillance in Animals

T. Kuiken,[1] F. A. Leighton,[2] R. A. M. Fouchier,[1] J. W. LeDuc,[3] J. S. M. Peiris,[4] A. Schudel,[5] K. Stöhr,[6] A. D. M. E. Osterhaus[1]*

Emerging infections have an enormous impact on public health, food supply, economies, and the environment (table S1). Human mortality from recently emerged diseases varies, ranging from less than 100 people thus far for H5N1 avian influenza to about 20 million for AIDS (1). Livestock production has been negatively affected by the direct mortality of animals from emerging infections and depopulation policies to protect the safety of international trade and to control the spread of pathogens. The environmental impact of emerging infections is of special concern for endangered wild animal populations, which can be pushed to the brink of extinction by such events (2). Animals, and particularly wild animals, are thought to be the source of more than 70% of all emerging infections (3) (see figure, A to F, right; table S1). Therefore, surveillance in animals for zoonotic pathogens—pathogens of nonhuman vertebrate animals that may be transmitted to humans under natural conditions—is critical for managing these infections (4).

Currently, pathogen surveillance in animals is usually the responsibility of government departments of agriculture. Its quality varies greatly among countries and typically does not include wildlife. Internationally, a list of pathogens that affect international trade, including many important zoonoses, are reported to the World Organization for Animal Health (OIE) by its member countries (5). The recently restructured Animal Health Information System of the OIE includes an International Early Warning System through which member countries have agreed to report immediately any and all of six different defined categories of animal disease occurrences, and an International Monitoring System through which absence, presence, or changes in status of diseases listed by the OIE are reported every 6 months and additional information is reported annually.

Other international organizations involved in pathogen surveillance of animals include the United Nations Food and Agriculture Organization (FAO) and the World Health Organization (WHO). The International Society for Infectious Diseases, a nongovernmental organization, maintains the Internet-based Program for Monitoring Emerging Diseases (Pro-MED-mail), which regularly carries reports of new epidemiological events around the world. Pathogen surveillance in wild animals, where and when it occurs at all, is without clear international reporting conventions. The few existing surveillance systems generally focus on a limited number of animal species and pathogens (6, 7). This information is collated annually from 167 countries by the OIE Wildlife Disease Working Group.

The current system does not provide a sufficient level of vigilance for several reasons. First, pathogen surveillance in domestic animals is generally confined to pathogens with known economic impacts. Second, pathogen surveillance in wildlife is less intensive to nonexistent, particularly in developing countries. Although some programs for wildlife exist, the number of species and geographical areas covered are very small compared with the many wild animal species (more than 15,000 species of mammals and birds alone) and their worldwide distribution. Finally

and most importantly, there is lack of integration among pathogen surveillance systems in humans, domestic animals, and wildlife. The sharp separation of the mandates of organizations charged with human and animal health management has encouraged total independence in their operations for a number of years. Below are some recent examples in which pathogen surveillance in animals, or its integration with public health, has faltered.

In November 2002, SARS first appeared in humans in southern China (8). The causative agent, a previously unknown coronavirus, was probably initially transmitted to humans by captive wild animals being sold as food (9) (see figure, C and D). No surveillance system was in place to detect pathogens in these marketed wild species. Thus, when the human pathogen was identi-
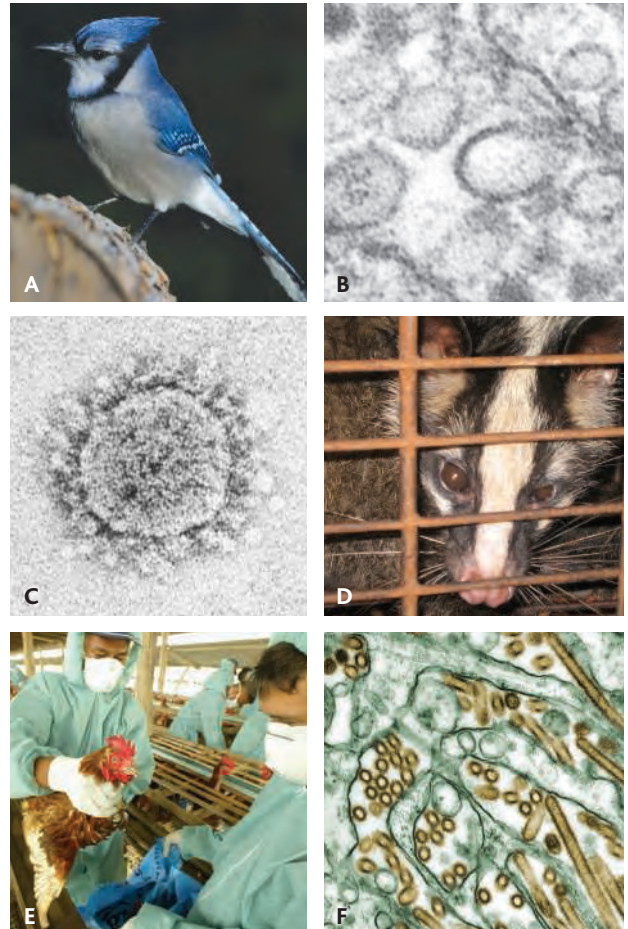
Emerging pathogens and their animal hosts. (**A**) A blue jay, one of the corvid species suffering high mortality from infection with West Nile virus, visualized (**B**) by transmission electron microscopy. (**C**) Negative contrast electron micrograph of SARS-associated coronavirus and (**D**) a masked palm civet, one of the species implicated in transmission of this virus to humans. (**E**) Culling of poultry infected with highly pathogenic avian influenza virus, subtype H5N1, shown (**F**) by transmission electron microscopy. Photo credits in (16).

[1]Department of Virology, Erasmus MC, 3015 GE Rotterdam, The Netherlands. [2]Canadian Cooperative Wildlife Health Centre, University of Saskatchewan, Saskatoon, SK, S7N 5B4, Canada. [3]National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA. [4]Department of Microbiology, University of Hong Kong, Hong Kong SAR, China. [5]World Organization for Animal Health (OIE), Paris, France. [6]World Health Organization, Department of Communicable Disease Surveillance and Response, Geneva, Switzerland.

*Author for correspondence. E-mail: a.osterhaus @erasmusmc.nl

fied, there was no infrastructure or historic information available to guide and speed identification of potential wild animal reservoirs. There were no archived tissue samples or epidemiological records available from game farms on which some of these animals were raised, from the national and international trade routes along which they were moved from wild populations to urban markets, or from the markets themselves (*10*).

In February 2003, an outbreak of highly pathogenic avian influenza, subtype H7N7, occurred in poultry in The Netherlands (*11*, *12*) and spread to Belgium and Germany. Some 30 million chickens were destroyed. H7N7 influenza virus infection also was confirmed in 89 people, with one fatality (*11*). Phylogenetic analysis showed that the new strain most likely had originated from free-living ducks and had evolved into a highly pathogenic variant after introduction into poultry farms (*11*). Subsequent serological screening of the Dutch poultry industry showed that H7 influenza virus had been affecting the Dutch poultry industry several months before the major epidemic (*12*), but that its presence had not been recognized.

Since December 2003, epidemic influenza due to a highly pathogenic H5N1 virus strain has devastated the poultry industry in southeastern Asia (*13*) (see figure, E and F). As of 27 July 2005, H5N1 virus infection, mainly due to direct transmission from birds, had been confirmed in 109 people, of whom 55 had died (*14*). Although highly pathogenic avian influenza is listed by OIE as a disease requiring immediate reporting, the actual reporting of the successive outbreaks in this epidemic by various countries was slow in several instances, with documented delays as long as 7 weeks between recognition and international reporting (*13*). These delays contributed to the failure of regional and international organizations to contain or control this epidemic. The delays in reporting are likely due in part to the fact that the reporting authorities in some countries have dual and conflicting mandates both to report disease occurrences and to foster their country's export status for animal products.

The main players in global health management are the WHO, FAO, and OIE. The interrelatedness of animal and human health in the 21st century is drawing these formerly independent organizations together as never before. The ongoing avian influenza outbreak in Asia's poultry, combined with the constant threat of a human influenza pandemic, has provided the momentum to join forces against this disease. This is exemplified by the recent formation of an OIE/FAO Network of Expertise on Avian Influenza (OFFLU), which will work in collaboration with the WHO

Influenza Network; the Global Strategy for the Progressive Control of Highly Pathogenic Avian Influenza, written by FAO and OIE in collaboration with WHO; and the FAO/OIE/WHO Consultation on Avian Influenza and Human Health, which proposed risk reduction measures for producing, marketing, and living with animals in Asia. Having demonstrated its feasibility for avian influenza, these international organizations should now expand this joint approach to zoonotic pathogens in general.

We recommend the creation of a joint expert working group either under the auspices of or with representation from the WHO, FAO, OIE, and other key stakeholders such as the World Conservation Union (IUCN). The mission of this expert working group should be to design and implement a global animal surveillance system for zoonotic pathogens that gives early warning of pathogen emergence, is closely integrated with public health surveillance, and provides opportunities to control such pathogens before they can affect human health, food supply, economies, or biodiversity. Canada's National Wildlife Disease Strategy, which aims to manage existing and emerging diseases that originate from wild animals, might serve as a useful template to develop such a system (*15*).

The tasks of the expert working group should include identifying the largest gaps in current surveillance systems in humans, domestic animals, and wildlife and determining the most cost-effective methods to fill them (e.g., sampling of animals at live animal markets, game farms, and ports of entry; establishment of systems for adequate preservation of samples from site of collection to the laboratory); identifying political barriers to effective surveillance (e.g., conflicting mandates of responsible authorities) and determining methods to deal with them appropriately (e.g., an international program of insurance against economic losses that may be triggered by rapid and transparent reporting); determining practical ways to integrate human and animal surveillance at the national and international level (including data management, communication networks, and regional laboratory networks); facilitating development and distribution of validated diagnostic tests for rapid and sensitive screening of samples for zoonotic pathogens in the full range of potential host species; and encouraging scientists and funding agencies to study the underlying mechanisms of disease emergence in order to find new methods to deal with this problem.

We estimate the annual start-up costs for such a working group at US$4 to $5 million for the first 3 years, based on a core group of 10 people with broad knowledge of

human, domestic animal, and wildlife health. Making use of the worldwide expertise in the relevant disciplines, this group would design and implement a global animal surveillance system for zoonotic pathogens. The costs of this system would depend in part on the risk that human society is willing to take, but would be only a fraction of the savings made by early detection and rapid response to emerging pathogens. Zoonotic pathogens do not respect national borders, so it is in the best interests of wealthy nations to invest in improved animal surveillance programs in all parts of the world.

Emerging diseases are a major challenge to the biological safety of the world in the 21st century. Renewed effort, resources, and scientific innovations, as well as new degrees of integration are required to meet this challenge. Linking comprehensive pathogen surveillance of wild and domestic animals—the principal source of emerging infections—with public health surveillance at the national and international level, and improving the quality and coverage of these surveillance programs, will make an important and essential contribution to the detection and control of emerging zoonotic infections.

References and Notes
1. Joint United Nations Programme on HIV/AIDS (UNAIDS), "2004 Report on the global AIDS epidemic: Fourth global report" (UNAIDS, Geneva, 2004).
2. P. D. Walsh *et al.*, *Nature* **422**, 611 (2003).
3. L. H. Taylor, S. M. Latham, M. E. Woolhouse, *Philos. Trans. R. Soc. London Ser. B* **356**, 983 (2001).
4. H. Kruse, A.-M. Kirkemo, K. Handeland, *Emerg. Infect. Dis.* **10**, 2067 (2004).
5. World Organization for Animal Health (OIE), *Terrestrial Animal Health Code* (OIE, Paris, ed. 13, 2004).
6. F. A. Leighton, *Rev. Sci. Tech. OIE* **14**, 819 (1995).
7. T. Mörner, D. L. Obendorf, M. Artois, M. H. Woodford, *Rev. Sci. Tech. OIE* **21**, 67 (2002).
8. J. S. M. Peiris, K. Y. Yuen, A. D. M. E. Osterhaus, K. Stöhr, *N. Engl. J. Med.* **349**, 2431 (2003).
9. Y. Guan *et al.*, *Science* **302**, 276 (2003).
10. D. Bell, S. Roberton, P. R. Hunter, *Philos. Trans. R. Soc. London Ser. B.* **359**, 1107 (2004).
11. R. A. Fouchier *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 1356 (2004).
12. G. Koch *et al.*, *Infect. Ziekten Bull.* **14**, 112 (2003).
13. OIE, "Update on avian influenza in animals in Asia (type H5)" 11 August 2005 (www.oie.int/downld/avian%20influenza/a_ai-asia.htm).
14. WHO, "Cumulative number of confirmed human cases of avian influenza A/(H5N1) reported to WHO," (www.who.int/csr/disease/avian_influenza/country/cases_table_2005_07_27/en/print.html).
15. Environment Canada, "Canada's National Wildlife Disease Strategy" 31 May 2004 (www.cws-scf.ec.gc.ca/cnwds/index_e.cfm).
16. Photo credits: (A), Town of Hudson, NH, USA; (B), (C), and (F), Centers for Disease Control and Prevention, Atlanta, GA, USA; (D), Hume Field, Department of Primary Industries and Fisheries, Brisbane QLD, Australia; (E), Australian Broadcasting Corporation, Sydney NSW, Australia.

PERSPECTIVES

IMMUNOLOGY

Dietary Factors and Immunological Consequences

Timothy Hla

What we eat is arguably linked to how robust our immunity is, but confusion still reigns as to the molecular details of how various nutrients and components in our diet can specifically affect our complex immune system. An exciting paper by Schwab *et al.* on page 1735 in this issue (*1*) shows that 2-acetyl-4-tetrahydroxy-butylimidazole (THI), a component of caramel food colorant III used in food products including beer and barbeque sauce, suppresses immunity by increasing the amount of a lipid called sphingosine 1-phosphate in tissues of our immune system. Sphingosine 1-phosphate has attracted much attention recently as a regulator of the immune and cardiovascular systems (*2*). This study is of considerable interest as it draws attention to dietary modulation of immune function by the regulation of lipid metabolism.

Many studies have focused on the pharmacological regulation of the cell surface receptors for sphingosine 1-phosphate [G protein–coupled receptors called $S1P_{1-5}$; (*3*)]. In contrast, our knowledge of sphingosine 1-phosphate metabolism is scant (*4*). The lipid is present in all eukaryotes, and although it is produced intracellularly during sphingolipid metabolism, it is also secreted by some of the cells that produce it. In mammals, a steep concentration gradient of sphingosine 1-phosphate exists between the vasculature and tissues, thereby exposing resident vascular and immune cells bearing a cognate receptor to excess ligand (*5*). As lymphocytes traffic in and out of the vascular system, they must navigate this steep sphingosine 1-phosphate gradient. Schwab *et al.* discovered that THI attenuates the activity of sphingosine 1-phosphate lyase in lymphoid tissues. This intracellular enzyme irreversibly degrades sphingosine 1-phosphate into phosphoethanolamine

and 2-hexadecanal. Thus, by neutralizing the concentration gradient of sphingosine 1-phosphate, THI perturbs T cell egress, causing a logjam in lymphoid tissue. Immunosuppression ensues (*1*).

Naïve lymphocytes circulate through lymphatic and vascular conduits (see the figure). In addition, they traffic in and out of peripheral tissues by crossing the vascular and lymphatic endothelial cell barriers.

Modulation of lymphocyte trafficking is a new approach to control immune pathology. Recent work with the experimental compound FTY720, which binds to the extracellular domain of sphingosine 1-phosphate receptors, has revealed an unexpected role of sphingosine 1-phosphate in the control of lymphocyte traffic (*6*). Expression of the sphingosine 1-phosphate receptor $S1P_1$ (formerly known as EDG-1) in T cells is critical for their proper egress from lymph nodes, Peyer's patches, and the thymus (*7*). This receptor is expressed on the surface of T cells when the extracellular concentration of sphingosine 1-phosphate is low; however, the receptor is rapidly internalized by cells upon exposure to the higher concentrations of the ligand (*8*). The concentration of sphingosine 1-phosphate is extremely high in the blood (0.4 to 1.5 µM)—about 25
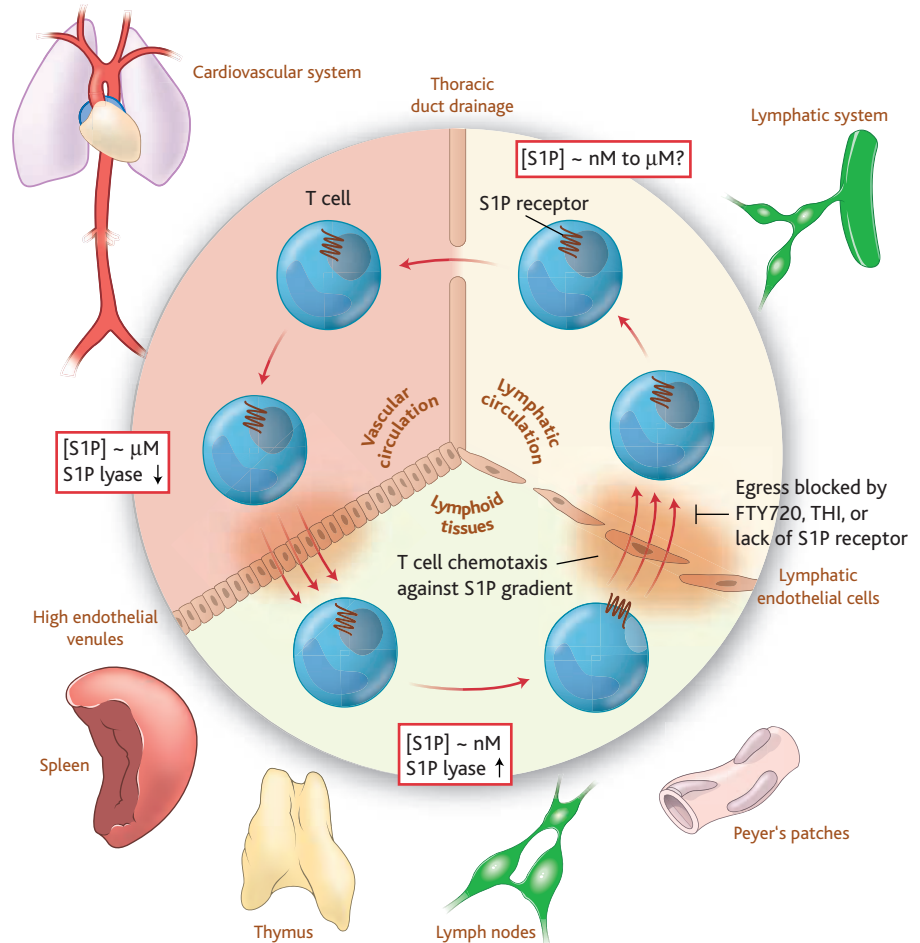
Sphingosine 1-phosphate gradients and T cell trafficking. In response to chemokines, T cells in the vascular system home into lymphoid tissues by traversing high endothelial venules. Low concentrations of sphingosine 1-phosphate ([S1P]) in the lymphoid tissues enables the expression of the sphingosine 1-phosphate receptor on the surface of T cells, hence allowing their chemotaxis response into the lymphatic system. This steep gradient of sphingosine 1-phosphate is maintained in part by the action of a sphingosine 1-phosphate lyase that degrades this lipid mediator. Inhibition of the lyase by dietary factors perturbs the gradient of sphingosine 1-phosphate and reduces circulating lymphocytes, resulting in immunosuppression.

The author is at the Center for Vascular Biology, Department of Cell Biology, University of Connecticut Health Center, Farmington, CT 06030, USA. E-mail: hla@nso2.uchc.edu

CREDIT: CHRIS BICKEL AND PRESTON HUEY/*SCIENCE*

times that in thymic tissue (*1*). Thus, in lymphoid tissues, the sphingosine 1-phosphate receptor is present on the surface of immune cells. But during transit in the blood and lymphatic fluid, the receptor is sequestered in the intracellular endosomal vesicles. The expression of sphingosine 1-phosphate receptors allows the cells to sense the direction of egress by moving against the vascular concentration gradient. FTY720 is effective in suppressing the immune response (mobilization of T cells into the vascular and lymphatic systems) because it persistently induces internalization of the sphingosine 1-phosphate receptor and presumably interferes with the compass system of the cell.

Schwab *et al.* observed that oral administration of THI to mice increased the concentration of sphingosine 1-phosphate in lymphoid organs. Consequently, this induced internalization of the receptor and accumulation of lymphocytes, resulting in lymphopenia and immunosuppression. Taking a clue from previous findings that the immuosuppressive effects of THI are reversed by vitamin B_6, the authors implicated sphingosine 1-phosphate lyase, which needs vitamin B_6 as a cofactor to degrade sphingosine 1-phosphate. The authors provide pharmacological evidence that THI attenuates enzyme activity in vivo, and further support the finding in mouse hematopoietic cells in which expression of the lyase is knocked down by RNA interference. Another inhibitor of vitamin B_6–dependent enzymes also acted as an immunosuppressant. These findings are of interest as they illuminate the role of sphingosine 1-phosphate lyase in immune regulation.

Sphingosine 1-phosphate lyase was originally discovered in the yeast *Saccharomyces cerevisiae* because of its ability to confer sphingosine tolerance (*9*). It is involved in many processes including resistance to the chemotherapeutic drug cisplatin, cell survival in mammals, embryonic development in *Drosophila melanogaster* and *Caenorhabditis elegans*, and the regulation of a transcription factor in flies (*4*). Because vitamin B_6 deficiency is associated with immune deficiency in humans, these findings suggest that additional dietary factors may modulate immunity by influencing sphingosine 1-phosphate metabolism.

The findings of Schwab *et al.* suggest that sphingosine 1-phosphate gradients play a fundamental role in the immune response. How this gradient is generated is not clear. But mammals appear to expend considerable energy maintaining the vascular gradient of this lipid. For example, vascular endothelial cells secrete sphingosine kinase to generate sphingosine 1-phosphate extracellularly (*10*). Other enzymes in the metabolic pathway, such as sphingosine 1-phos-phate phosphatase, and genetic manipulation of the lyase should be examined to further elucidate how the gradient is generated and maintained. Expression of sphingosine 1-phosphate receptors is inducible by growth factors and cytokines, but receptor internalization may limit the cellular response. The concept of a sphingosine 1-phosphate gradient may also help resolve an apparent discrepancy in the biology of this lipid mediator—it is abundantly present in blood under normal conditions, yet is capable of inducing numerous pathologies including inflammation, angiogenesis, and cell proliferation (*2*). The hydrophobic as well as amphipathic physicochemical property of sphingosine 1-phosphate may afford unique modes of receptor activation. Indeed, lipid gradients may behave differently than water soluble polypeptide gradients in the context of signaling within tissues.

On a more practical level, modulation of activity and expression of sphingosine 1-phosphate lyase by intrinsic and dietary factors may lead to immunological consequences. It is not yet known how THI inhibits the lyase. Whether sphingosine 1-phosphate lyase is involved in regulation of the vascular homeostasis and/or growth also remains to be determined. But this work provides the molecular basis for the regulation of immunity by dietary factors—lending credence to the saying, we are what we eat.

References and Notes
1. S. R. Schwab *et al.*, *Science* **309**, 1735 (2005).
2. T. Hla, *Semin. Cell Dev. Biol.* **15**, 513 (2004).
3. J. Chun *et al.*, *Pharmacol. Rev.* **54**, 265 (2002).
4. J. D. Saba, T. Hla, *Circ. Res.* **94**, 724 (2004).
5. Y. Igarashi, Y. Yatomi, *Acta Biochim. Pol.* **45**, 299 (1998).
6. V. Brinkmann, J. G. Cyster, T. Hla, *Am. J. Transplant.* **4**, 1019 (2004).
7. M. L. Allende, J. L. Dreier, S. Mandala, R. L. Proia, *J. Biol. Chem.* **279**, 15396 (2004).
8. C. H. Liu *et al.*, *Mol. Biol. Cell* **10**, 1179 (1999).
9. J. D. Saba, F. Nara, A. Bielawska, S. Garrett, Y. A. Hannun, *J. Biol. Chem.* **272**, 26087 (1997).
10. N. Ancellin *et al.*, *J. Biol. Chem.* **277**, 6667 (2002).
11. This work is supported by NIH grants HL67330 and HL70694.

MATERIALS SCIENCE

Nanosprings Take Shape

Brian A. Korgel

Nanoscale coiled springs—in particular, piezoelectric nanosprings—have potential applications in nano-electromechanical systems (NEMS) as actuators, resonators, and sensors. Piezoelectric materials are mechanically responsive to external electric fields, and an applied voltage can be used to stretch or compress the spring with physical distortions on the order of nanometers, making them ideal for NEMS devices. However, fabrication of crystalline coiled springs with nanometer-scale dimensions is a daunting task. NEMS fabrication relies on lithographic patterning (by optical or electron beam or by physical imprint) of photoresist to shape the material components. These techniques, combined with etching and deposition methods, are adequate for producing structurally simple components such as the micrometer-thick polysilicon beams that are widely used as NEMS springs. Unfortunately, these methods are not suitable for coiled springs, because conventional lithographic techniques are inherently two-dimensional. Furthermore, even if a template could be fabricated by lithography, it is no small task to deposit a single-crystal material in the shape of a coiled spring. Therefore, alternative "bottom-up" fabrication processes have been explored that yield structurally complex materials like nanosprings without relying on lithographic patterning. On page 1700 in this issue, Gao and co-workers report an important breakthrough in bottom-up synthesis of nanosprings made from zinc oxide (*1*).

In the years preceding this advance, two relatively successful bottom-up approaches to nanospring fabrication have emerged: (i) glancing angle deposition (GLAD) (*2*, *3*) and (ii) spontaneous nanowire coiling in vapor-liquid-solid (VLS) growth or vapor-solid (VS) growth (*4–10*). GLAD is a physical vapor deposition process performed with the source material located at a specific angle with respect to a rotating substrate. The technique has successfully produced high-quality arrays of coiled metal and ceramic nanowires with controlled pitch and helicity (that is, left- or right-handedness). The GLAD technique, however, is limited to the formation of dense arrays with relatively large wire diameters (>100 nm) and only amorphous ceramics. Spontaneous nanowire coiling during VLS or VS growth, on the other hand, can give rise to nanocoils with narrower diameters (<100 nm in most cases), yet the coiling process is only partially understood for most materials systems, and

The author is in the Department of Chemical Engineering, Texas Materials Institute, Center for Nano- and Molecular Science and Technology, The University of Texas at Austin, Austin, TX 78712, USA. E-mail: korgel@mail.che.utexas.edu
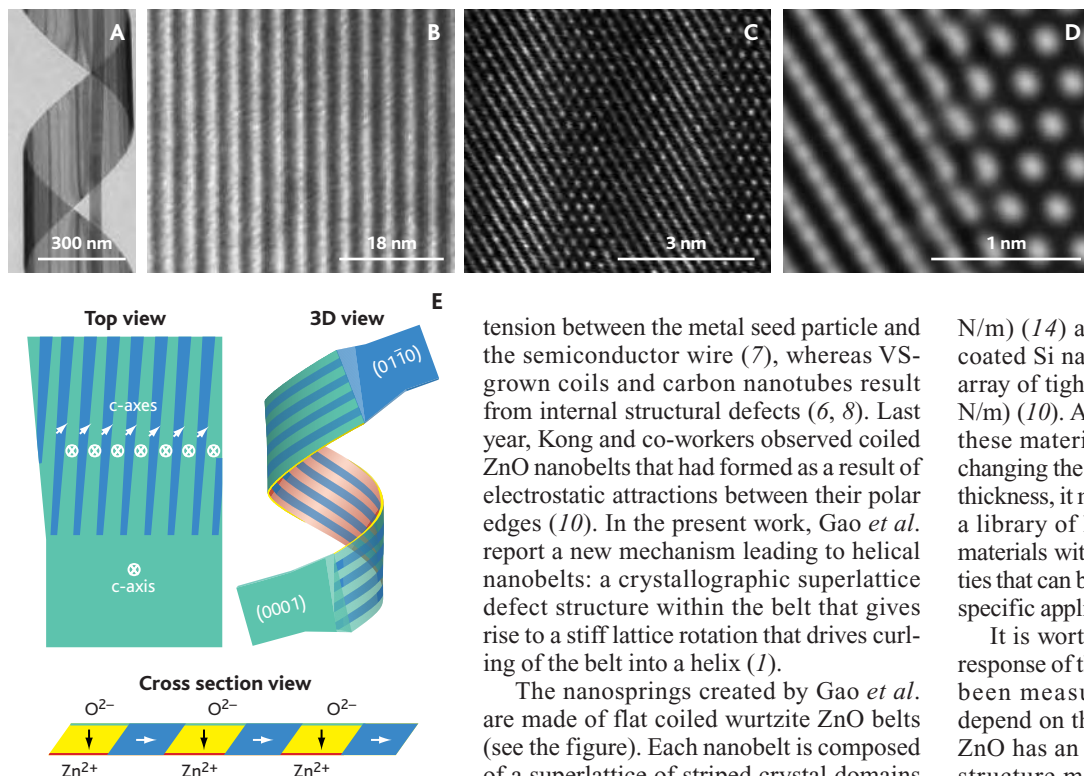
Spring growth. (**A**) Scanning electron micrograph of a ZnO nanospring. (**B** to **D**) Transmission electron microscopy images at incasing magnification reveal a superlattice structure of crystalline "stripes" with alternating crystallographic direction. As shown in the schematic (**E**), the alternating layers in the superlattice structure give rise to a rigid crystal twist that leads to the helical structure. [Adapted from Gao *et al.* (*1*)]

in most cases the yield of coils is extremely low. Gao and co-workers (*1*) have now obtained crystalline ZnO nanosprings by VS growth in abundances of up to 10% in the synthesized material. Once formed, the nanosprings were manipulated to measure the mechanical spring constant.

These results (*1*) highlight the increasingly successful efforts by many researchers to develop reproducible chemical strategies to fabricate coiled nanoscale springs. Some of the first nanospring materials— helical carbon fibers and tubes—were observed a few decades ago (*4*). Many researchers since have observed coiled semiconductor nanowires and carbon nanotubes produced under various VLS and VS reaction conditions, including helical nanowires of carbon filaments (*4*), amorphous silicon nitride (*5*), polycrystalline Si (*6*), boron carbide (*7*), silicon carbide (*8*), silicon oxide (*9*), and crystalline ZnO (*1*, *10*). Although the yields are typically very low and far from commercial viability, some materials design criteria have begun to emerge: In VLS-grown coils, it appears that the helical shape relates to the surface

tension between the metal seed particle and the semiconductor wire (*7*), whereas VS-grown coils and carbon nanotubes result from internal structural defects (*6*, *8*). Last year, Kong and co-workers observed coiled ZnO nanobelts that had formed as a result of electrostatic attractions between their polar edges (*10*). In the present work, Gao *et al.* report a new mechanism leading to helical nanobelts: a crystallographic superlattice defect structure within the belt that gives rise to a stiff lattice rotation that drives curling of the belt into a helix (*1*).

The nanosprings created by Gao *et al.* are made of flat coiled wurtzite ZnO belts (see the figure). Each nanobelt is composed of a superlattice of striped crystal domains ~3.5 nm wide with crystallographic orientations that alternate growth direction between $[1\bar{1}00]$ and $[2\bar{1}\bar{1}2]$. Each crystal domain stripe is oriented with its *c* axis nearly perpendicular to one another (offset by about 4°). The 4° offset in growth direction of each crystalline "stripe" gives rise to a rigid lattice rotation that leads to the bending of the ribbon into a helix.

Extended defects in semiconductor nanowires and nanobelts have not been extensively studied; however, they have been observed in several nanowire materials, including ZnO (*9*), GaAs, GaP, Ge, and Si nanowires (*11*, *12*). In most cases, extended defects appear to have arisen from subtle perturbations in the growth conditions. The superlattice structure in the ZnO nanobelts, however, is believed to result from the electrostatic charge on the polar surfaces, which is dissipated by the transformation of the polar surfaces into nonpolar ones by a rotation in crystal symmetry. The small lattice mismatch between the superlattice stripes gives rise to only a very small strain, but the slight lattice rotation leads to the coiling of the nanobelt. Unlike the helical ZnO nanobelts reported by Kong *et al.* (*10*), the coils are rigid and relatively inflexible. A small belt segment cut from the nanospring retained its curvature.

Recent advances in nanomanipulators and nanomechanical testing have made it fairly routine to pick a nanospring out of a pile of fabricated material to perform mechanical testing (*13*, *14*). A spring with an applied force, *F*, and reasonably small deformation,

Δ, follows Hooke's Law: $F = k\Delta$. The spring constant *k* of the superlattice ZnO nanohelices was found to be ~4 N/m, which corresponds to an elastic modulus of ~50 GPa. The superlattice ZnO nanosprings are stiffer than carbon nanocoils (0.12 N/m) (*13*) and coiled multi-wall carbon nanotubes (0.7 N/m) (*14*) and more flexible than metal-coated Si nanocoils (8.75 N/m) (*3*) and an array of tightly packed SiO nanocoils (590 N/m) (*10*). Although the spring constants of these materials are potentially tunable by changing the pitch, spring diameter, and wire thickness, it may be more practical to develop a library of known obtainable nanospring materials with measured mechanical properties that can be called upon when needed for a specific application.

It is worth noting that the piezoelectric response of the ZnO nanosprings has not yet been measured, but it should strongly depend on the crystal structure of the belt. ZnO has an anisotropic hexagonal crystal structure made up of O^{2-} and Zn^{2+} ions, which are aligned into electric dipoles in the crystallographic *c* direction that give rise to its piezoelectricity. The electric dipoles respond to an applied voltage and lead to a crystal distortion. The piezoelectric response will therefore depend strongly on the nanospring crystallinity, including the details of crystallographic growth direction, helical angle, and the nature of the extended defects. The lattice direction mismatch in the ZnO superlattice may give rise to different piezoelectric properties than single-crystal ZnO nanobelts. It will be exciting to see the results of these kinds of experiments as the fabrication methods are further advanced toward applications.

References
1. P. X. Gao *et al.*, *Science* **309**, 1700 (2005).
2. M. W. Seto, K. Robbie, D. Vick, M. J. Brett, L. Kuhn, *J. Vac. Sci. Technol. B* **17**, 2172 (1999).
3. J. P. Singh *et al.*, *Appl. Phys. Lett.* **84**, 3657 (2004).
4. R. T. K. Baker, P. S. Harris, S. Terry, *Nature* **253**, 37 (1975).
5. S. Motojima, S. Ueno, T. Hattori, K. Goto, *Appl. Phys. Lett.* **54**, 1001 (1989).
6. Y. H. Tang *et al.*, *J. Appl. Phys.* **85**, 7981 (1999).
7. D. N. McIlroy, D. Zhang, Y. Kranov, M. G. Norton, *Appl. Phys. Lett.* **79**, 1540 (2001).
8. H.-F. Zhang, C.-M. Wang, L.-S. Wang, *Nano Lett.* **2**, 941 (2002).
9. H.-F. Zhang, C.-M. Wang, E. C. Buck, L.-S. Wang, *Nano Lett.* **3**, 577 (2003).
10. X. Y. Kong, Y. Ding, R. Yang, Z. L. Wang, *Science* **303**, 1348 (2004).
11. F. M. Davidson, R. Wiacek, B. A. Korgel, *Chem. Mater.* **17**, 230 (2005).
12. T. Hanrath, B. A. Korgel, *Small* **1**, 717 (2005).
13. X. Chen *et al.*, *Nano Lett.* **3**, 1299 (2003).
14. M. A. Poggi *et al.*, *Nano Lett.* **4**, 1009 (2004).

p53 and PUMA: A Deadly Duo

Karen H. Vousden

Despite its importance in protecting us from malignancies, the specific activities of the p53 protein that allow it to function as a tumor suppressor have been hard to reconcile. A paper by Chipuk *et al.* on page 1732 of this issue (*1*) now promises to reduce the complexity of the p53 response to cellular stresses by bringing together two seemingly separate activities of p53—one nuclear and one cytoplasmic—into a single, unified model.

The ability of p53 to function as a transcription factor is generally considered to be its main physiological property (*2*). Expression array technology has revealed long lists of potential gene targets of p53, with the accompanying problem of sorting the wheat from the chaff. Analyses of cells defective in some of these p53-target genes have established several of them as important mediators of the p53 response. For example, p21Waf1/Cip1, an inhibitor of an enzyme (a cyclin-dependent kinase) that controls the cell division cycle, is key to the proliferative arrest induced by p53. More recent studies have indicated a surprisingly strong dependence of p53-mediated apoptosis on the presence of PUMA, a protein with clear apoptotic potential (*3*). Combined with compelling evidence that the ability to function as a transcription factor is essential for the antineoplastic activity of p53, it is tempting to conclude that regulation of gene expression is the key to its suppressive effects on tumorigenesis.

However, complicating this simple model have been persistent indications that p53 also functions through nonnuclear, transcriptionally independent mechanisms (*4*). Early studies showing that p53 fragments lacking part of its DNA binding domain could still induce cell death, and that p53-mediated death was not necessarily dependent on new protein synthesis, elicited some concern as to their physiological relevance. However, more recent evidence has revealed a function for p53 outside the nucleus—as a binding partner of the anti-apoptotic/prosurvival members of the the Bcl2 superfamily such as Bcl2 and Bcl-xL (*5*) (see the figure, top). In this respect p53 seems to function like some of the proapoptotic members of the Bcl2

superfamily—a group of proteins defined by the presence of a region of sequence similarity called the BH3 domain—and so-called BH3-only proteins. This activity of p53 as a BH3-only protein comes as quite a surprise, because there is no obvious BH3 domain within p53. But PUMA, a validated p53-induced apoptotic protein, clearly contains a BH3 domain. This presents a conundrum: Why would p53 activate expression of PUMA, a BH3-only protein, if p53 functions as a BH3-only protein itself?

Stress
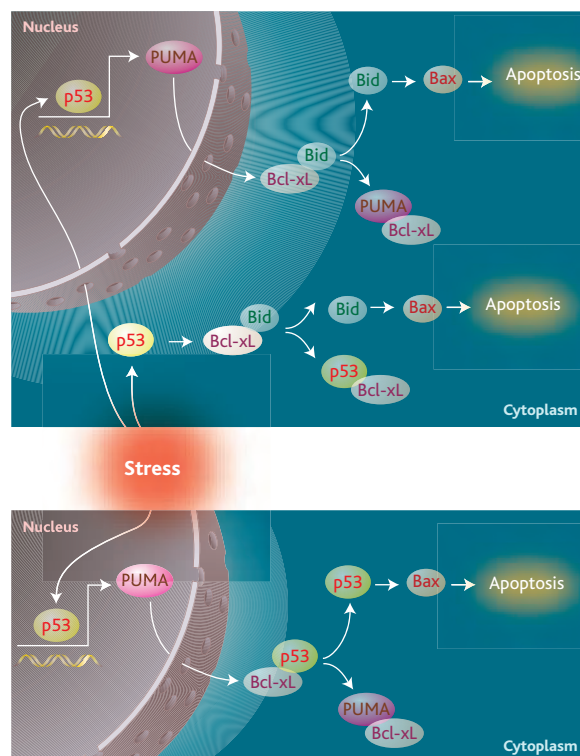
Models for the apoptotic activity of p53. (Top) Apoptotic functions of p53 include a nuclear role as a transcription factor that activates expression of gene targets including PUMA, and a nonnuclear role in the cytoplasm and mitochondria, involving interaction with anti-apoptotic members of the Bcl2 family of proteins. Each of these may result in mitochondrial outer membrane permeabilization and cell death. A number of cytoplasmic and mitochondrial activities of p53 have been proposed (see text). In this example, mitochondrial p53 releases Bid from Bcl-xL. In these models, the nuclear, cytoplasmic, and mitochondrial functions of p53 are not directly linked, although they may cooperate to fully activate cell death. **(Bottom)** The transcriptional and cytoplasmic functions of p53 are brought together by Chipuk *et al.* (*1*) in the same pathway where p53 activates the expression of PUMA, which then serves to release cytoplasmic p53 from the inhibitory interaction with Bcl-xL. The released p53 is then free to directly activate Bax. In this model p53 acts at two steps in the same chain of events, leading to mitochondrial outer membrane permeabilization and apoptosis.

The answer to this riddle appears to lie in the detail of how BH3-only proteins themselves function. The end point of the chain of events induced by BH3-only proteins is activation of the proapoptic proteins Bax and/or Bak (also members of the Bcl2 superfamily) to allow mitochondrial outer membrane permeabilization and release of apoptogenic factors such as cytochrome c. This then triggers the activation of a cascade of proteolytic enzymes called caspases and the apoptotic demise of the cell. But BH3-only proteins come in two flavors—activators and enablers (*6*, *7*). Activator BH3-only proteins function by transiently binding to, and directly activating, Bax and Bak to start the apoptotic ball rolling at the mitochondria. The enabler BH3-only proteins take a more elaborate and indirect path to activate Bax and Bak by forming a complex with the anti-apoptotic proteins such as Bcl2 or Bcl-xL. In the current model of these pathways, the binding of enabler BH3-only proteins to the anti-apoptotic proteins by itself has no real consequences. Rather, the importance of these interactions is that they displace activator BH3-only proteins (or even Bax or Bak themselves) from an inhibitory interaction with the anti-apoptotic proteins, so releasing them to drive the initial steps of the apoptotic cascade. The beauty and simplicity of the Chipuk *et al.* paper lies in uniting the evidence that PUMA is an enabler (*8*) and p53 an activator (*9*) into a model in which PUMA functions to release p53 from Bcl-xL, thereby freeing p53 to activate Bax (see the figure, bottom). In the absence of cell stress, only low levels of p53 are expressed. Thus, the nuclear level of p53 is insufficient to drive transcriptional activation of PUMA and the small amount of cytoplasmic p53 is held inactive by Bcl-xL. An unexpected stressful "jolt" to the cell, such as DNA damage or oncogene activation, results in a rapid increase of nuclear p53 and transcriptional activation of PUMA expression. PUMA then binds to Bcl-xL, hence freeing p53 to activate Bax.

Of course, there are other models of p53 function that might be considered, and

The author is at the Beatson Institute for Cancer Research, Garscube Estate, Switchback Road, Bearsden, Glasgow G61 1BD, UK. E-mail: k.vousden@beatson.gla.ac.uk

alternative explanations are necessary to accommodate all the available data. A number of studies have shown that p53 moves to the mitochondria in response to stress, suggesting that translocation and binding of p53 to mitochondrial Bcl2 and Bcl-xL may also trigger apoptosis (*10*). In this model the mitochondrial p53 could function as an enabler BH3-only protein to release activators like Bid (see the figure, top). Mitochondrial p53 can also show activator functions such as binding to Bak, which results in the release of Bak from Mcl1 (an anti-apoptotic protein similar to Bcl2 and Bcl-xL) (*11*). Direct activation of Bax by p53 appears to take place in the cytosol (*9*). To address the relative importance of enabler and activator functions of p53, Chipuk *et al.* used a Bcl-xL mutant that binds p53 but not PUMA. Although both wild-type and mutant Bcl-xL inhibited p53-mediated apoptosis, expression of PUMA could only reverse the effect of wild-type Bcl-xL. The implication is therefore that p53 has to be released from Bcl-xL by PUMA to induce apoptosis and that the binding of p53 to Bcl-xL is by itself not a proapoptotic signal. While providing elegant support for the activator function of p53, this observation does not preclude a function for p53 as an enabler and overall it seems likely that coordination of the nuclear, cytoplasmic, and mitochondrial functions of p53 will contribute to the ultimate response to stress.

A number of questions arise out of the Chipuk *et al.* study, including whether p53 requires PUMA, or PUMA requires p53, to induce cell death. The answer to the first question, at least in some cell types, seems to be yes. Deletion of PUMA by genetic knockout or knockdown by RNA interference strategies strongly impairs p53-dependent apoptosis in certain cell systems (*3*). But PUMA may not be unique in this function. A number of other BH3-only proteins, such as Noxa, are transcriptionally activated by p53 and play an essential role in the p53 apoptotic response in some cell types. So it seems likely that under certain conditions, proteins like Noxa might substitute for PUMA. Less clear is whether PUMA might require p53. Initial studies have indicated that PUMA expression is enhanced by withdrawal of serum from cells or by inducing stress in the endoplasmic reticulum, and that this induction of PUMA is p53 independent (*12*, *13*). Furthermore, like p53, the transcription factor E2F1 can activate PUMA expression, and in this case PUMA was shown to contribute to apoptosis without requiring p53 (*14*). Although it is extremely exciting to consider p53 as a functional homolog of an activator BH3-only protein, other proteins, such as Bim and Bid, also display this activity. Because PUMA has an extremely high affinity for the anti-apoptotic Bcl2-like proteins (*15*), it seems reasonable to suppose that in addition to

releasing p53, PUMA will also release any bound Bim or Bid. Indeed, the study by Chipuk *et al.* suggests that this is possible, and may go some way to explaining the strong apoptotic activity seen following PUMA expression in some p53-null cells. But regardless of these details, which will undoubtedly be the subject of strong debate and intense research, the p53-PUMA relationship suggested by this model provides a very satisfying explanation to the quandary of why p53 should have evolved both transcriptional and cytoplasmic functions.

References
1. J. E. Chipuk *et al.*, *Science* **309**, 1732 (2005).
2. B. Vogelstein, D. Lane, A. J. Levine, *Nature* **408**, 307 (2000).
3. J. Yu, L. Zhang, *Cancer Cell* **4**, 248 (2003).
4. K. S. Yee, K. H. Vousden, *Carcinogenesis* **26**, 1317 (2005).
5. S. Erster, U. M. Moll, *Biochem. Biophys. Res. Commun.* **331**, 843 (2005).
6. A. Letai *et al.*, *Cancer Cell* **2**, 183 (2002).
7. T. Chittenden, *Cancer Cell* **2**, 165 (2002).
8. T. Kuwana *et al.*, *Mol. Cell* **17**, 525 (2005).
9. J. E. Chipuk *et al.*, *Science* **303**, 1010 (2004).
10. M. Mihara *et al.*, *Mol. Cell* **11**, 577 (2003).
11. J. I. Leu, P. Dumont, M. Hafey, M. P. Murphy, D. L. George, *Nat. Cell Biol.* **6**, 443 (2004).
12. J. Han *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11318 (2001).
13. C. Reimertz, D. Kogel, A. Rami, T. Chittenden, J. H. Prehn, *J. Cell Biol.* **162**, 587 (2003).
14. T. Hershko, D. Ginsberg, *J. Biol. Chem.* **279**, 8627 (2004).
15. L. Chen *et al.*, *Mol. Cell* **17**, 393 (2005).

BIOPHYSICS

Harvesting Energy by Improving the Economy of Human Walking

Arthur D. Kuo

Humans are tremendously flexible when converting almost any food into energy, but they are both inflexible and insatiable in their demand for mobile energy in another form: electricity. In the developed world, many people seem more concerned with their cell phone battery life than with their next meal. Given the plentiful and, in many cases, increasing supply of stored onboard energy (that is, fat), could humans not generate the necessary electricity themselves? Hand-operated generators are both inexpensive and effective for short-term use. But a less distracting alternative for the long term might be to generate elec-

tricity from walking, given that walking is how humans already expend much of their daily energy. On page 1725 in this issue, Rome *et al.* (*1*) describe a new backpack device that harvests far more energy from locomotion than other methods of obtaining energy from walking, while costing the wearer a surprisingly low amount of metabolic energy. It works by extracting energy through an oscillating sprung mass. Why it works so well is unclear. Perhaps the device reduces the mechanical work required of muscles to walk while carrying a load.

The energy-harvesting backpack succeeds with an approach different from other attempts. More common is to generate electricity from the compression of a shoe (*2*), for example, with piezoelectric crystals. Shoe placement is logical, because it allows the wearer to apply an entire body weight to

the device. But displacement must be very small to avoid disrupting gait. The result so far is that low power is generated on average, less than 1 W electrical. An alternative is to place the device in parallel with the limbs, with displacement provided by gross limb motion. Such devices are ungainly and require the human to produce extra force. The approach of the energy-harvesting backpack (see the figure) is entirely different. Conceptually, it resembles the self-winding mechanism of an automatic wristwatch, where power is generated from an oscillating payload, excited inertially through the wearer's motion. Neither force nor displacement is imposed; both arise from the device's dynamics. This idea would seem to have little merit at a larger scale. The watch mechanism is useful because its inertia is small and power requirements are minuscule. A 29-kg wristwatch would hardly be tolerable. Yet the backpack is both comfortable and effective, generating nearly 4 W of electrical power with a similar payload.

The energy-harvesting backpack is also a curiosity because of how well it works. It costs metabolic energy both to walk while carrying a fixed payload (about 590 W) and to perform work on a generator (about 48 W metabolic

The author is in the Departments of Mechanical Engineering and Biomedical Engineering, University of Michigan, Ann Arbor, MI 48109–2125, USA. E-mail: artkuo@umich.edu
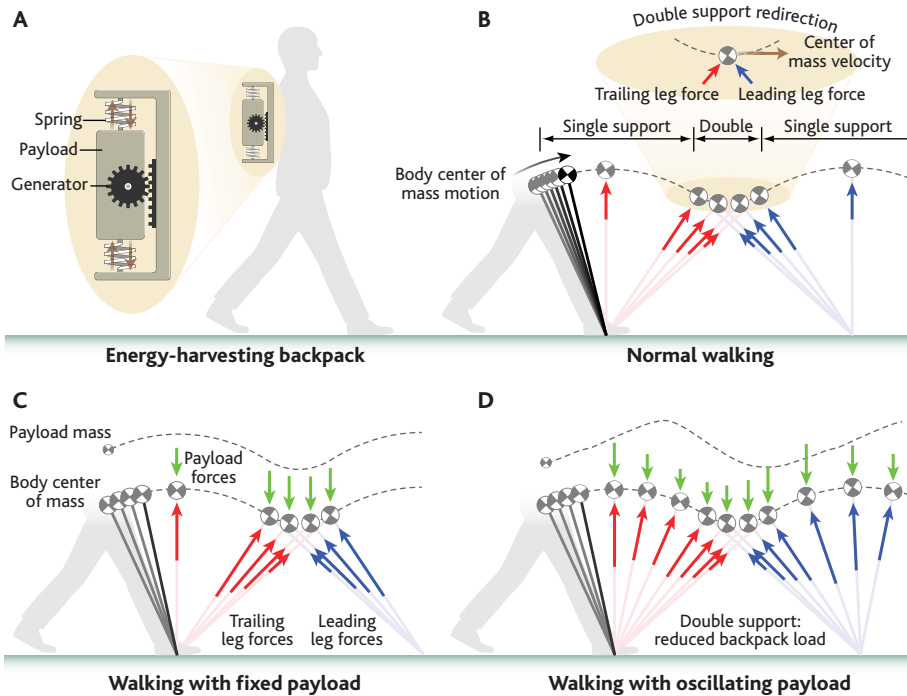
Power walking. Simple models of an energy-harvesting backpack and its relation to human walking. (**A**) Conceptual schematic of backpack, where the payload (the mass to be carried) is supported by springs, and electrical energy is generated when the payload moves up and down. (**B** to **D**) Schematics demonstrate hypothesis for how forces are exerted on the body's center of mass by the legs and backpack. In normal walking (**B**), the body is supported alternately by one and then two legs (single and double support). The single support leg is relatively straight and the center of mass moves like an inverted pendulum, with no need for energy input. During double support, the legs exert forces to redirect the center of mass along a U-shaped trajectory between successive pendular arcs. The trailing leg performs positive work and the leading leg negative work, with the two nearly canceling each other despite costing metabolic energy. (**C**) The backpack's payload is fixed to its frame and exerts additional forces on the body center of mass. The inverted pendulum of single support remains energy-conservative, but additional positive and negative work is performed during double support, costing more energy. (**D**) The payload is unlocked and oscillates vertically. If the motion is properly phased, the backpack load acting on the center of mass may be slightly less than in (C) during double support, requiring less redirection work. The load must then be slightly greater during single support, but if the leg is relatively straight it can support extra load with little muscle force. Some additional muscle work is needed as input to the generator, but this is partially offset by the savings in redirection work. This may explain how the backpack in (D) can generate power more economically than would be expected from (C).

for 12 W mechanical work, due to muscle's efficiency of about 25%). But with the energy-harvesting backpack, subjects expended less than the sum of these two costs, saving nearly 30 W over the expected metabolic rate. It is unlikely that muscle's fundamental ability to perform work had improved. Perhaps the backpack reduces the amount of work needed. But where must work be performed? Walking consists of alternating phases where one leg is kept relatively straight and acts like an inverted pendulum (3, 4). This allows body weight to be supported with relatively little effort, and for the body's center of mass to freely travel along an arc. Some energy must realistically be expended during these pendulum-like phases, but not enough to explain the overall energetic cost of walking, let alone the mysterious advantage of the energy-harvesting backpack.

The explanation may lie in the transition between pendulum-like walking steps, when the body's center of mass is redirected from one pendular arc to the next (5, 6). The center of mass is located near the hip joints and undergoes a small U-shaped displacement during this step-to-step transition, which occurs mainly when both legs contact the ground. Force is exerted by, and directed along, each leg, with the leading leg performing negative work on the center of mass and the trailing leg positive work. The leading leg's force is at such an angle with the direction of center of mass displacement that negative work is unavoidable, if the center of mass is to be redirected to another pendular arc. This negative work is thought to be largely dissipated as an energy loss. An equal magnitude of positive work performed by the trailing leg cancels

this loss, as is needed to walk at steady speed. Positive and (to a lesser degree) negative work both cost positive metabolic energy, contributing substantially to the overall cost of normal walking. Now consider that the addition of a fixed payload to the body's mass increases the work required for the step-to-step transition. But a sprung payload, oscillating vertically, exerts a fluctuating load on the body. If the energetic cost of the pendulum-like phase is relatively insensitive to additional load (because the leg need not perform work) and the step-to-step transition is relatively more sensitive (because both legs are performing work), then the sprung payload can prove advantageous by exerting more downward force during the pendulum phase than the step-to-step transition. This might reduce both the work required of step-to-step transitions and the peak forces exerted by the backpack, simultaneously improving energy consumption and comfort. The actual phasing of the prototype device's motion relative to the wearer, not measured here, could indicate whether this hypothesis explains the backpack's advantage.

The energy-harvesting backpack is novel because it generates useful amounts of electrical power while costing less metabolic energy than would be expected. The saving only applies in comparison to a person already walking with a heavy load, but that same person might also be eager to avoid carrying a set of batteries. The backpack also highlights an important aspect of human walking, indicating that muscles perform work that cancels mechanically but costs metabolically. The backpack's sprung payload may reduce this work. Present understanding is incomplete, but there is no obvious reason why the backpack cannot be improved to reduce muscle work requirements still further (7). One could then generate electricity while carrying a load more economically and with greater comfort than with a conventional backpack. Future backpackers might be less concerned about both cell phone battery life and their next meal.

References

1. L. C. Rome, L. Flynn, E. M. Goldman, T. D. Yoo, *Science* **309**, 1725 (2005).
2. J. A. Paradiso, T. Starner, *IEEE Pervasive Computing* **1**, 18 (2005); available online at www.media.mit.edu/resenv/pubs/papers/2005-02-E-Harvesting PervasivePprnt.pdf.
3. R. M. Alexander, *Science* **308**, 58 (2005).
4. R. M. Alexander, *Appl. Mech. Rev.* **48**, 461 (1995).
5. J. M. Donelan, R. Kram, A. D. Kuo, *J. Exp. Biol.* **205**, 3717 (2002).
6. A. D. Kuo, J. M. Donelan, A. Ruina, *Exerc. Sport Sci. Rev.* **33**, 88 (2005).
7. A. Chatterjee, R. Pratap, C. K. Reddy, A. Ruina, *Int. J. Robotics Res.* **21**, 621 (2002).

Magnetic Domain-Wall Logic

D. A. Allwood,[1] G. Xiong,[2] C. C. Faulkner,[3] D. Atkinson,[2] D. Petit,[3] R. P. Cowburn[3,4]*

"Spintronics," in which both the spin and charge of electrons are used for logic and memory operations, promises an alternate route to traditional semiconductor electronics. A complete logic architecture can be constructed, which uses planar magnetic wires that are less than a micrometer in width. Logical NOT, logical AND, signal fan-out, and signal cross-over elements each have a simple geometric design, and they can be integrated together into one circuit. An additional element for data input allows information to be written to domain-wall logic circuits.

Conventional microelectronic integrated circuits (ICs) work by controlling the flow of electrons through transistor switches. Digital signals are represented in the IC by the presence or absence of electrical charge (Table 1). Electrons can offer more, however. In addition to electrical charge, electrons also possess the quantum mechanical property of spin. Unlike electrical charge, however, spin can have two directions, conventionally known as "up" and "down," permitting alternate representations of binary digits. For example, the magnetization of a small ferromagnetic element is the classical limit of electron spin and has long been used to store information in magnetic recording. During the last decade, a number of researchers have been building the new technology of spintronics, in which the spin and the charge of the electron are used in microelectronic ICs to represent bits and carry out data processing. In general, this technique promises lower-powered, higher-speed, nonvolatile devices with which to build the next generation of computing technology (1, 2).

Spintronics development has followed different approaches in the semiconductor and magnetism communities. The semiconductor approach involves creating and manipulating spin-polarized electrons in a semiconductor host, with information being represented as either spin "up" or spin "down." At the moment, the lack of a suitable room-temperature ferromagnetic semiconductor has limited the development of functioning devices (3), although much progress has been made in understanding the manipulation of spin in semiconductors by using optical probes (4).

The magnetism community has taken a different tack. After the discovery of giant magnetoresistance exhibited by metallic ferromagnetic/nonmagnetic multilayers (5), researchers have developed a large number of room-temperature devices by using ferromagnetic metals such as nickel, iron, and cobalt (1). Information is represented in these devices by the direction of magnetization in a small ferromagnetic element. One such device, called a magnetic tunnel junction (MTJ), forms the building block of magnetic random-access memory (MRAM), which is a nonvolatile, high-density, high-speed memory technology close to commercialization (6). Furthermore, recent demonstrations of the spin-momentum transfer effect (7–11), in which magnetization is operated on directly by an applied current, have added further heat to this already-exciting field by providing a new interface mechanism between the worlds of electronics and magnetism. We and other researchers have been working to extend magnetic nonvolatile memory to develop the required elements for viable magnetic logic technologies.

Magnetic logic. Digital microelectronics is a combination of memory and logic. The basic Boolean logic functions, such as AND, NOT, and XOR, allow the digital IC to combine numbers from memory in arithmetic calculations. MRAM is manufactured using a complementary metal-oxide semiconductor (CMOS)–compatible process and will, therefore, indirectly affect microelectronic logic by allowing large amounts of high-speed, high-density, nonvolatile memory to be embedded with the semiconductor microprocessor. The emerging field of magnetic logic seeks to redesign the principles of operation of microelectronic logic at the lowest level to make direct use of ferromagnetism.

Attempts have been made (12–14) to implement magnetic logic based on a single-electron transistor architecture designed by the University of Notre Dame (15). These schemes, called magnetic cellular automata, use a network of magnetostatically coupled magnetic elements. Information propagates by magnetic solitons running across the lattice of interacting magnetic elements, and logic functions are performed by summing stray magnetic fields at nodal dots that have well-defined switching thresholds. One of the challenges with such schemes is that the magnetostatic interaction field between ferromagnetic elements is usually weaker than the demagnetizing field within the element; consequently, any physical defects in the shape of the magnetic elements tend to block the propagation of information, and the device becomes excessively intolerant to fabrication faults.

A number of MTJ-based magnetic logic schemes have been proposed. In one class of these (16–18), information enters into the logic gate by the currents in multiple bit lines. The so-called "free layer" of the MTJ will rotate into the direction of the net magnetic field from the combined currents, effectively acting as a nonlinear summing element. This in turn changes the resistance of the junction, which can be used to control the current in subsequent bit lines. Such a scheme has a number of advantages, including that the devices are based on existing MTJ technology and that the logic function can be programmed by changing the magnetization direction of the reference layer in the MTJ. This makes MTJs attractive for field-programmable gate arrays (19), where many different applications use the same hardware; the precise hardware function is defined by programming a configuration of the storage elements. Perhaps even more appealing is the prospect of on-the-fly reconfigurability, because the function-defining magnetic hard layer can be reversed in nanoseconds, allowing hardware to adaptively track the optimized architecture for the computation in hand (1). A disadvantage of these schemes is that high current densities must be switched every time the data change, a process which requires high magnetoresistance ratios and large transistors. A variation on this theme exists, in which an MTJ is used to bias a conventional electronic logic gate (20). In this case, the MTJ is only used to define the logic function; the actual computation is performed entirely in classical electronics.

As an alternative, we have developed a magnetic logic architecture referred to as "domain-wall logic," which uses no transistors and exhibits very little heating caused by data switching. A domain wall is a mobile interface between regions of oppositely aligned magne-

[1]Department of Engineering Materials, University of Sheffield, Sir Robert Hadfield Building, Portobello Street, Sheffield S1 3JD, UK. [2]Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK. [3]Blackett Physics Laboratory, Imperial College London, Prince Consort Road, London SW7 2BW, UK. [4]Durham Magneto Optics Ltd., The University of Durham, Durham DH1 3HP, UK.

*To whom correspondence should be addressed. E-mail: r.cowburn@imperial.ac.uk

tization. In particular, a submicrometer planar nanowire made from a soft magnetic material such as Permalloy ($Ni_{80}Fe_{20}$) has been shown to form an excellent conduit for domain walls (21–23). The high shape anisotropy of the nanowire ensures that magnetization prefers to align with the long axis of the wire. These two possible directions form the basis of the binary information representation (Table 1), with a magnetic domain wall acting as the transition edge in a changing signal. Domain walls can be propagated through complex networks of nanowires under the action of an externally applied magnetic field. This field rotates in the plane of the device and acts as both the clock and the power supply. Previous work (24) has shown how a cusp-shaped planar nanowire can be used to reverse the direction of magnetization. Given that we define a logical "1" as the magnetization pointing in the direction of propagation of domain walls and a logical "0" as the magnetization opposing the direction of propagation of domain walls, the cusp in effect performs the logical NOT operation. In order to implement any arbitrary logic function, it is necessary to have some additional basis functions. At least one two-input function, such as AND or OR, is required to complement the NOT function so that any computational calculation can be performed. Two routing functions are also required for most complex logic circuits: a fan-out structure, which makes two identical copies of an input signal; and a cross-over structure, which allows two signals to pass over each other without interference. It is also necessary that the applied magnetic field requirements of the functions should be mutually compatible, so that a single global rotating magnetic field can be applied to the entire circuit, with all of the different functions operating together. We have found that this can be achieved for an architecture consisting of logical NOT, logical AND, fan-out, and cross-over junctions (Table 1). Furthermore, we have developed a field-addressable data-input element for providing logic circuits with data on which to operate. We describe the development of these devices below.

Domain-wall logic elements. Figure 1A shows a magnetic logic circuit made up of a NOT-gate cusp, a fan-out junction, and a cross-over junction (25). For measurement, the entire circuit was placed in a rotating magnetic field. As shown previously (24, 26), the NOT gate was fabricated in a loop structure to ensure that at least one domain wall existed in the loop and to enable easy experimental testing. What is new here, however, is that a domain wall propagating around the loop in Fig. 1A must also pass through a cross-over structure, as indicated schematically in Fig. 1B. To obtain a well-defined starting state of a single domain wall before measurement, the device magnetization was first saturated in a large (>200 Oe) magnetic field before adjacent pairs of domain walls were annihilated by using a low-amplitude
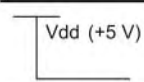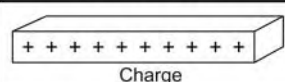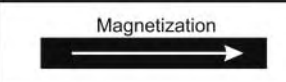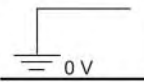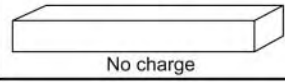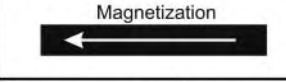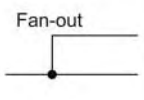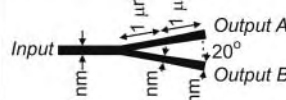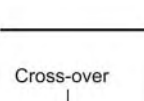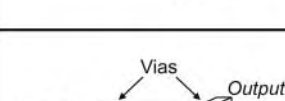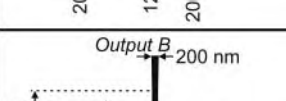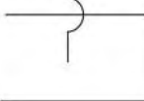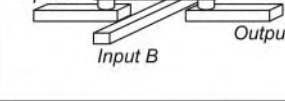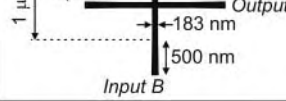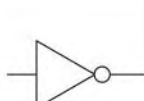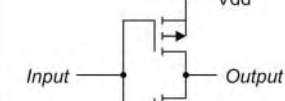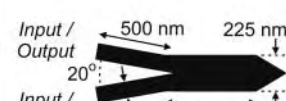
rotating field, as before (26). Because of the synchronous nature of the logic device, circuit geometry defined the domain-wall propagation time around the NOT gate/loop to be 0.5 field cycles through the NOT gate and 1 field cycle for each 360° loop. Hence, a single domain wall round trip should take 2.5 field cycles, leading to a 5–field cycle magnetization switching period. The fan-out element formed part of this loop but did not affect the domain wall round-trip time. One fan-out output fed back into the loop, whereas the other extended into a long arm to provide a monitor of the loop magnetization.

Within a suitable counterclockwise rotating field, magneto-optical Kerr effect (MOKE) magnetometry (24, 27, 28) measurement from the position marked by an asterisk in Fig. 1A in-

dicated a switching period of 5 field cycles (Fig. 1C), confirming that the NOT gate, fan-out, and cross-over elements were working correctly.

Of the four logic architecture elements, the cross-over junction is the most challenging to achieve experimentally, and its operation is highly sensitive to the nanowire dimensions. To pass across the cross-over element, a domain wall must expand fully across the junction, an energetically costly process, before being able to propagate further along the output wire. However, domain walls must not be allowed to propagate along the orthogonal wire direction, or else the digital information in the structure will have been altered. In contrast, a domain wall propagating through a fan-out junction will gradually expand from the input wire as the junction widens, before

Table 1. Symbols for electronic logic, together with the appropriate CMOS circuit element and a schematic drawing of the optimized domain-wall logic element (dimensions shown refer to the design rule used here of 200-nm–wire width connecting logic elements and 1-μm turning radius of corners). The fan-out, cross-over, and logical AND junctions all contain tapered regions to connect to the 200-nm-wide wire links. Vdd is the supply voltage.

dividing into two separate walls once the output wires are reached.

The integration of all four logic elements was completed in the nanowire network shown in Fig. 2A, consisting of a NOT gate, an AND gate, two fan-out junctions, and one cross-over junction. Our previous work (29) has shown that the switching field of the AND-gate output wire depends on whether neither, one, or both of the input wires contains a domain wall, with the switching field reducing with an increasing number of incident domain walls. In order to achieve logical AND functionality, the AND gate is operated within an elliptical rotating magnetic field with a dc field bias H_x^{DC} in the direction of H_x. This is similar to how previous pseudo-AND operations have been achieved in other single-layer magnetic systems (12, 30). The remain-

der of the magnetic circuit in Fig. 2A was designed to sequentially supply the AND gate with all four possible logical input combinations for a two-input device. However, all wire junctions must be able to tolerate the dc field bias, because this is applied globally.

A NOT gate within a feedback loop was used as a signal generator for the rest of the network and should have a 3–field cycle switching period (24, 26). A fan-out element fed domain walls out of the loop and to a second successive fan-out element, where the domain walls were divided into two paths again. MOKE measurement at position I in Fig. 2A indicates a switching period of 3 field cycles (Fig. 2B, trace I), confirming that the NOT gate and sequential fan-out elements were working correctly. Between positions I and II (Fig. 2A), we can infer that domain walls will be delayed by 0.5 field cycles (Fig. 2B, trace II). However, for domain walls from the second fan-out junction to reach position III, they must pass through an additional loop created by the inclusion of the cross-over junction. The magnetization at position III, therefore, will be delayed by 1 field cycle, compared with that at position II (Fig. 2B, trace III). The magnetization direction at positions II and III determines the logical input state of the AND gate. An AND gate has an output value of 1 only when both inputs are 1, and is 0 for all other conditions. The measurement of position IV of the magnetic circuit (Fig. 2B, trace IV) showed this to be the case (using the convention here that a high MOKE signal refers to logical 1), demonstrating that the AND gate was operating correctly together with the other three element types.

Data input. Any useful circuit must be able to receive data from the outside world. Figure 3A shows a circular shift register structure containing eight NOT gates and one fan-out junction, where one of the NOT gates has an enlarged central stub region to lower the magnetization reversal field and make this a data-input element. The stray magnetic field from a current-carrying conductor could be used to write magnetic data directly and locally into the enlarged stub (31, 32). However, in this case we have designed the stub such that field amplitudes required to write data to it lie within the operation range of the other NOT gates and fan-out junction. We can therefore write data by modulating the amplitude of the globally applied rotating magnetic field. The rotating field thus acts simultaneously as a power supply, clock, and serial data channel.

Figure 3, B to E, shows the operating principle of a data-input element within a clockwise rotating field. From the initial magnetization state (Fig. 3B), a large amplitude field H_x^{write} nucleates a domain wall in the element that propagates through the NOT-gate junction and divides into domain walls 1 and 2 along the input/output wires (Fig. 3C). Domain wall 1 is routed around a corner of the same handedness as the applied field rotation and will continue to propagate around the shift register (Fig. 3D). In contrast, domain wall 2 is initially routed around a corner of opposite handedness to the applied field rotation. As the field rotates further, domain wall 2 must reverse its direction and pass back through the NOT gate (Fig. 3D). At the NOT gate, the returning domain wall must again split into two. One
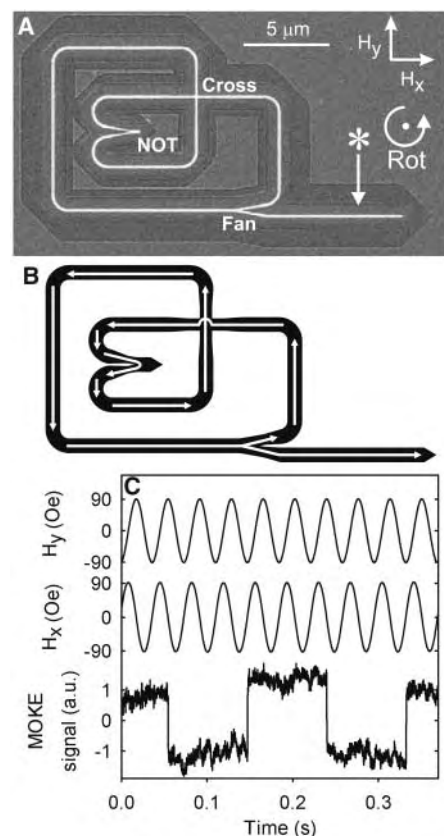

Fig. 1. (A) Focused ion beam (FIB) image of a magnetic nanowire loop containing a NOT gate, fan-out junction, and cross-over junction. Only the bright lines indicate the presence of magnetic material; all other features are artifacts of the fabrication process. The directions of rotating field components (H_x and H_y), and the sense of field rotation (Rot) are indicated, and the asterisk denotes the position of MOKE measurement. **(B)** Schematic diagram with arrows illustrating the route of a domain wall propagating through the magnetic structure within a counterclockwise rotating field. **(C)** MOKE trace obtained from the position marked with an asterisk of the nanowire structure within an applied counterclockwise rotating field with amplitudes $H_x^0 = H_y^0 = 91$ Oe.

Fig. 2. (A) FIB image of a magnetic nanowire network containing one NOT gate, one AND gate, two fan-out junctions, and one cross-over junction. MOKE measurements were made at positions I and IV, indicated by asterisks, and positions II and III denote the inputs to the AND gate. Also indicated are the directions of field components (H_x and H_y) and the sense of field rotation (Rot). **(B)** MOKE traces describing the operation of the magnetic circuit within a counterclockwise rotating field with amplitudes $H_x^0 = 75$ Oe and $H_y^0 = 88$ Oe and dc offset of $H_x^{DC} = -5$ Oe. Experimental MOKE measurements from positions I and IV of the circuit are shown. Traces II and III are inferred from trace I and show the magnetization state of the AND gate's input wires.

part will propagate along what is the output arm of the NOT-gate junction (Fig. 3E), following original domain wall 1 with a half-cycle delay, whereas another part will propagate back along the data-input element to restore the initial magnetization state (Fig. 3E). It is important that this returning domain wall annihilates when it reaches the element end to avoid an oscillation condition. The output of the data-input element from this field sequence with a single nucleation event is, therefore, a pair of domain walls.

The "write" field pattern shown in Fig. 3F was only applied once to fill the shift register with a 5-bit data stream, with only the section between the dotted lines performing data writing (the final field cycle is used to ensure that all domain walls correctly enter the shift register loop). Before the writing process, all domain walls were annihilated by applying a low-amplitude rotating field (26). For writing a single data bit, the amplitude components of one half-cycle of field are $H_x^{write} = 138$ Oe and

$H_y^0 = 50$ Oe (directions defined in Fig. 3A), causing the magnetic data-input element to switch as described above. The field conditions for not writing data are $H_x^{no-write} = 90$ Oe and $H_y^0 = 50$ Oe. MOKE measurement with $H_x^0 = 90$ Oe and $H_y^0 = 50$ Oe of the shift register in its initialized configuration verified that no domain walls were present (Fig. 3G, trace I). After a single application of the "write" field pattern, the shift register contains several two–domain-wall packets that represent a binary data stream of "11010" (Fig. 3G, trace II). Because of the data writing procedure, a low-to-high transition in the MOKE signal corresponds to logical 1, whereas no transition across a full field cycle corresponds to logical 0. This data stream corresponds perfectly to the data stream in the "write" field pattern (Fig. 3F) and confirms the principle of data-input element operation described above. A delay of 1 hour between writing and reading returned the correct bit sequence, demonstrating the intrinsic nonvolatility of the shift register.
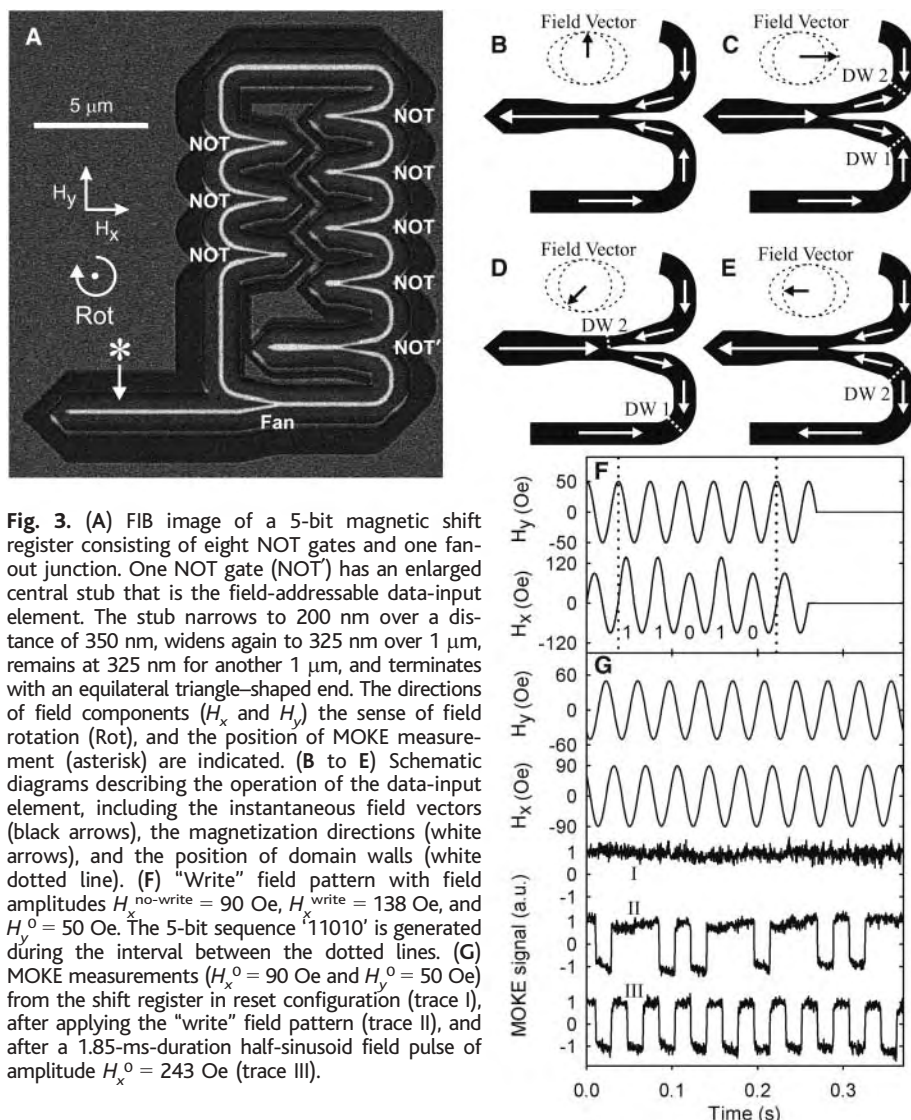
However, the ultimate room-temperature storage time will far exceed 1 hour, and wire width and thickness can be engineered to ensure that data retention times exceed 10 years. All information within the entire shift register can be removed by the application of a single bulk erase half-sinusoid field pulse of amplitude $H_x^0 = 243$ Oe and 1.85-ms pulse length, immediately filling the shift register with ten domain walls (Fig. 3G, trace III), regardless of the initial magnetization arrangement.

Outlook. We have thus demonstrated the four basis functions (NOT, AND, fan-out, and cross-over) operating simultaneously under the action of a single global applied rotating magnetic field. In principle, any logic circuit can now be implemented, simply by increasingly complex combinations of the four basis functions. All the demonstrations here used a 27-Hz magnetic field frequency to allow easy MOKE measurement and rapid device prototyping. Other initial experiments indicate that NOT gates operate correctly at low-kHz frequencies, with no upper limit observed. Furthermore, previous measurements of domain-wall propagation velocities in excess of 1000 m s^{-1} (23, 33) suggest propagation delays between wire junctions equal to 0.1 to 2 ns, depending on the design rule size. However, complex gyromagnetic behavior at wire corners (23) and domain-wall pinning at wire junctions are likely to increase these propagation delays and are the subject of ongoing investigation. Domain-wall dynamics are likely to be affected by future miniaturization, but this also remains untested. Additional work is required to integrate either the magnetic data-input element described above or current-carrying wires for data input with the four nanowire logic junctions. Furthermore, we are currently developing magnetoresistive logic read-out elements by using multilayer regions.

One of the most attractive features of domain-wall logic is its great simplicity (Table 1). Logical NAND is achieved in CMOS by using four transistors, whereas domain-wall logic uses two elements (NOT and AND). A logical AND function, usually requiring six CMOS transistors, can be achieved simply by bringing two magnetic nanowires together. Unlike Si CMOS architecture, a domain wall cross-over junction can be achieved in a single plane and without multilevel metallization, which means that in principle, extremely low-cost devices could be produced. Whereas most applications of magnetic logic (and indeed the wider field of spintronics) will involve a hybrid system on a chip that includes silicon-based CMOS, certain applications such as biomedical implants or wearable computing hardware would benefit from the ability to fabricate devices on, for example, flexible polyimide substrates. One could imagine nanowires constructed into three-dimensional (3D) neural networks or hugely dense 3D non-

Fig. 3. (**A**) FIB image of a 5-bit magnetic shift register consisting of eight NOT gates and one fan-out junction. One NOT gate (NOT') has an enlarged central stub that is the field-addressable data-input element. The stub narrows to 200 nm over a distance of 350 nm, widens again to 325 nm over 1 μm, remains at 325 nm for another 1 μm, and terminates with an equilateral triangle–shaped end. The directions of field components (H_x and H_y) the sense of field rotation (Rot), and the position of MOKE measurement (asterisk) are indicated. (**B** to **E**) Schematic diagrams describing the operation of the data-input element, including the instantaneous field vectors (black arrows), the magnetization directions (white arrows), and the position of domain walls (white dotted line). (**F**) "Write" field pattern with field amplitudes $H_x^{no-write} = 90$ Oe, $H_x^{write} = 138$ Oe, and $H_y^0 = 50$ Oe. The 5-bit sequence '11010' is generated during the interval between the dotted lines. (**G**) MOKE measurements ($H_x^0 = 90$ Oe and $H_y^0 = 50$ Oe) from the shift register in reset configuration (trace I), after applying the "write" field pattern (trace II), and after a 1.85-ms-duration half-sinusoid field pulse of amplitude $H_x^0 = 243$ Oe (trace III).

volatile memories. The ability to supply power, clock, master reset, and serial input with a single externally applied magnetic field, as demonstrated in Fig. 3, is particularly attractive in the 3D case, where signal access is limited. There may also be potential for interfacing domain-wall logic with emerging dilute ferromagnetic semiconductors that allow electrical control and sensing of magnetization (34).

The energy dissipated in each domain-wall logic element must, on very general thermodynamic grounds, be less than $2M_sH_aV$ per gate output transition, where M_s is the saturation magnetization of the magnetic material (800 electromagnetic units cm^{-3} for Permalloy), H_a is the amplitude of the externally applied magnetic field, and V is the volume of magnetic material switching in the gate and its output nanowire. For the devices described here, typical values for the energy per operation are 10^{-5} pJ ($2000k_BT$ at room temperature, where k_B is Boltzmann's constant and T is temperature), compared with the typical energy per gate of CMOS of 10^{-2} pJ for a 200-nm minimum feature size (35). This could allow large 3D domain-wall logic circuits to operate without overheating. However, domain-wall logic will not necessarily be a low–power-consumption technology, because there are considerable inefficiencies in generating magnetic fields. Absolute power consumption might be kept low by using small-area (few hundred μm^2) domain wall–CMOS hybrid devices, with strip lines for local field generation. However, domain-wall propagation by spin transfer (36–38) may ultimately overcome these inefficiencies altogether.

The future scaling performance of domain-wall logic depends on the interplay between thermodynamic stability and the required magnitude of the externally applied field. Both of these depend on F, the width of the nanowires that form the logic elements. If the width and thickness of the nanowires are scaled together, then the shape anisotropy remains unchanged. To first order, the strength of the externally applied magnetic field required to overcome fabricational edge roughness and discontinuities at the logic elements therefore remains constant. In this case, the energy per gate transition scales with the volume of magnetic material, proportional to F^3. Thus, at $F = 70$ nm, a 1.8-nm-thick magnetic logic device should dissipate 3×10^{-7} pJ ($70k_BT$ at room temperature), which is the lower limit for thermodynamic stability. For further reductions in F, the device thickness should be increased by scaling by $F^{-1/2}$ in order to keep the energy per gate transition unchanged. This will cause shape anisotropy to increase, and hence the required externally applied field for domain-wall propagation through wires and the structural discontinuities associated with wire junctions will increase as well. As with MRAM, the ultimate limit of scaling will be when the required externally applied field becomes impractically large (39). What is not currently known is how well defined some of the fine features in the logic elements, such as the point in the center of the NOT gate, have to remain as F is scaled down. This will determine the relationship between F and the required resolution of the lithography, and hence the relationship with Moore's law. We would expect that fine definition should become less important as F is reduced, because the exchange interaction will increasingly average the local magnetization response.

References and Notes

1. G. A. Prinz, *Science* **282**, 1660 (1998).
2. S. A. Wolf *et al.*, *Science* **294**, 1488 (2001).
3. T. Dietl, H. Ohno, F. Matsukura, J. Cibert, D. Ferrand, *Science* **287**, 1019 (2000).
4. Y. Ohno *et al.*, *Nature* **402**, 790 (1999).
5. M. N. Baibich *et al.*, *Phys. Rev. Lett.* **61**, 2472 (1988).
6. G. Grynkenich *et al.*, *MRS Bull.* **29**, 818 (2004).
7. J. C. Slonczewski, *J. Magn. Magn. Mater.* **159**, L1 (1996).
8. L. Berger, *Phys. Rev. B* **54**, 9353 (1996).
9. J. A. Katine, F. J. Albert, R. A. Buhrman, E. B. Myers, D. C. Ralph, *Phys. Rev. Lett.* **84**, 3149 (2000).
10. S. I. Kiselev *et al.*, *Nature* **425**, 380 (2003).
11. W. H. Rippard, M. R. Pufall, S. Kaka, S. E. Russek, T. J. Silva, *Phys. Rev. Lett.* **92**, 027201 (2004).
12. R. P. Cowburn, M. E. Welland, *Science* **287**, 1466 (2000).
13. G. Csaba, W. Porod, A. I. Csurgay, *Int. J. Circ. Theor. Appl.* **31**, 67 (2003).
14. A. Imre, G. Csaba, V. Metlushko, G. H. Bernstein, W. Porod, *Physica E (Amsterdam)* **19**, 240 (2003).
15. I. Amlani *et al.*, *Science* **284**, 289 (1999).
16. G. Reiss *et al.*, *Phys. Status Soldi* **201**, 1628 (2004).
17. A. Ney, C. Pampuch, R. Koch, K. H. Ploog, *Nature* **425**, 485 (2003).
18. R. Richter *et al.*, *Solid State Electron.* **46**, 639 (2002).
19. Z. Navabi, *Digital Design and Implementation with Field Programmable Devices* (Kluwer Academic Publishers, Dordrecht, Netherlands, 2005).
20. W. C. Black, B. Das, *J. Appl. Phys.* **87**, 6674 (2000).
21. R. D. McMichael, M. J. Donahue, *IEEE Trans. Magn.* **33**, 4167 (1997).
22. T. Ono *et al.*, *Science* **284**, 468 (1999).
23. D. Atkinson *et al.*, *Nat. Mater.* **2**, 85 (2003).
24. D. A. Allwood *et al.*, *Science* **296**, 2003 (2002).
25. All structures presented here are fabricated from thermally evaporated, 5-nm-thick Permalloy films by focused ion beam milling (24).
26. D. A. Allwood *et al.*, *J. Appl. Phys.* **95**, 8264 (2004).
27. D. A. Allwood, G. Xiong, M. D. Cooke, R. P. Cowburn, *J. Phys. D* **36**, 2175 (2003).
28. The high-sensitivity MOKE magnetometer (24, 26) was used with at least 1 min of signal averaging and a field rotation frequency of 27 Hz. The magnetometer can be used to select regions of complex magnetic nanowire networks for analysis with ~5 μm spatial resolution.
29. C. C. Faulkner *et al.*, *IEEE Trans. Magn.* **39**, 2860 (2003).
30. D. A. Allwood *et al.*, *Appl. Phys. Lett.* **81**, 4005 (2002).
31. A. Himeno *et al.*, *J. Appl. Phys.* **93**, 8430 (2003).
32. D. Atkinson, R. P. Cowburn, *J. Magn. Magn. Mater.* **290-291**, 165 (2005).
33. Y. Nakatani, A. Thiaville, J. Miltat, *Nat. Mater.* **2**, 521 (2003).
34. D. Chiba, M. Yamanouchi, F. Matsukura, H. Ohno, *Science* **301**, 943 (2003).
35. R. Waser, Ed., *Nanoelectronics and Information Technology* (Wiley VCH, Weinheim, Germany, 2003), p. 330.
36. M. Tsoi, R. E. Fontana, S. S. P. Parkin, *Appl. Phys. Lett.* **83**, 2617 (2003).
37. C. K. Lim *et al.*, *Appl. Phys. Lett.* **84**, 2820 (2004).
38. N. Vernier, D. A. Allwood, D. Atkinson, M. D. Cooke, R. P. Cowburn, *Europhys. Lett.* **65**, 526 (2004).
39. R. P. Cowburn, *J. Appl. Phys.* **93**, 9310 (2003).
40. This work was supported by Ingenia Holdings Ltd. D.A.A. and D.A. acknowledge the Engineering and Physical Sciences Research Council for Advanced Research Fellowship nos. GR/T02942/01 and GR/S51288/01, respectively. The work and results reported in this publication were obtained with research funding from the European Community under the Sixth Framework Programme Contract Number 510993: MAGLOG. The views expressed are solely those of the authors, and the other Contractors and/or the European Community cannot be held liable for any use that may be made of the information contained herein.

A Human-Specific Gene in Microglia

Toshiyuki Hayakawa,[1,2]* **Takashi Angata,**[1,2]*† **Amanda L. Lewis,**[1,3]
Tarjei S. Mikkelsen,[6] **Nissi M. Varki,**[1,4] **Ajit Varki**[1,2,5]‡

Recent studies have shown differences between humans and chimpanzees in sialic acid (Sia) biology, including Sia-binding receptors called Siglecs [1–3]. Analysis of the ~3-kb region of homology between human *SIGLEC11* (*hSIGLEC11*) and pseudogene *hSIGLECP16* (Fig. 1A) showed that an ~2-kb segment (designated A/A′) including the first five exons is 99.3% identical. However, the rest of this region (designated B/B′) has a much lower identity (94.6%). A phylogenetic tree of B/B′ from the human (h) and chimpanzee (c) genomes shows the topology expected from the gene orthology (Fig. 1B). In contrast, the A/A′ tree shows a within-species clustering of *hSIGLEC11* and *hSIGLECP16* (Fig. 1C), suggesting a recent gene conversion event.

Pairwise genetic distances between human and chimpanzee A/A′ regions were calculated, after excluding intron 2 and exons [4] (fig. S1), and compared with the standard human-chimpanzee genetic distance, calculated from downstream intronic regions (0.0141 ± 0.0013). The genetic distance between *hSIGLEC11* and *hSIGLECP16* is much smaller (0.0049 ± 0.0024), and that between corresponding chimpanzee forms is much larger (0.0337 ± 0.0065). Thus, although the presumed gene duplica-

tion event yielding ancestral *SIGLEC11* and *SIGLECP16* predated the human-chimpanzee common ancestor, the subsequent gene conversion occurred only in the human lineage. Both *hSIGLEC11* and *hSIGLECP16* are more closely related to *cSIGLECP16* than to *cSIGLEC11*, and the genetic distance between *hSIGLEC11* and *cSIGLECP16* is nearly the same as that between *SIGLECP16* orthologs [4]. Thus, *hSIGLECP16* converted *hSIGLEC11*. Inclusion of bonobo, gorilla, and orangutan *SIGLEC11* sequences confirmed that this gene conversion occurred only in the human lineage [4] (fig. S2). A search of hu-

man single-nucleotide polymorphism databases [4] suggests that it is universal to modern humans.

The converted region includes a 5′ upstream region and exons encoding the Sia recognition domain. We therefore studied the Sia-binding properties of recombinant, soluble Siglec-11 protein. Chimpanzee Siglec-11 showed more robust binding than hSiglec-11 (Fig. 2A), especially to Neu5Gc, the Sia type missing in humans [3]. Human Siglec-11 still bound oligosialic acids [$(Neu5Ac\alpha2-8)_{2-3}$], which are enriched in the brain.

Human brain cortex microglia showed strong Siglec-11 staining [5] in eight individuals. However, despite positive macrophage staining in other tissues, microglia showed only occasional staining in five chimpanzees (Fig. 2B) and none in two orangutans. Thus, brain microglia gained specific and prominent Siglec-11 expression in the human lineage, possibly due to changes of regulatory sequences in the 5′ upstream region.

This work also supports the notion that pseudogenes are better called "potogenes" or potential genes [6]. We suggest that this human-specific gene conversion event may be related to the evolution of genus *Homo*. Multiple approaches could shed light on its evolutionary significance, including further analysis of the converted region, promoter studies, calculation of the event timing, detection of Siglec-11 ligands in the brain, and a search for humans with *SIGLEC11* mutations.

References and Notes

1. A. Varki, *Nat. Genet.* **36**, 1034 (2004).
2. Chimpanzee Sequencing and Analysis Consortium, *Nature*, in press.
3. A. Varki, T. Angata, *Glycobiology*, in press.
4. Materials and methods are available as supporting material on *Science Online*.
5. T. Angata *et al.*, *J. Biol. Chem.* **277**, 24466 (2002).
6. E. S. Balakirev, F. J. Ayala, *Annu. Rev. Genet.* **37**, 123 (2003).
7. Supported by NIH grant no. R01GM32373 (A.V.), a Japan Society for the Promotion of Science Fellowship for Research Abroad (T.H.), and by the Mathers Foundation (A.V.).

Fig. 1. Gene structure and phylogenetic analyses. (**A**) Comparison of *SIGLEC11* and the similar region of *SIGLECP16*. Exons are represented by solid and open boxes. (**B** and **C**) Phylogenetic relationships of (B) B/B′ and (C) A/A′. The label at the internode represents bootstrap support for 1000 replications.
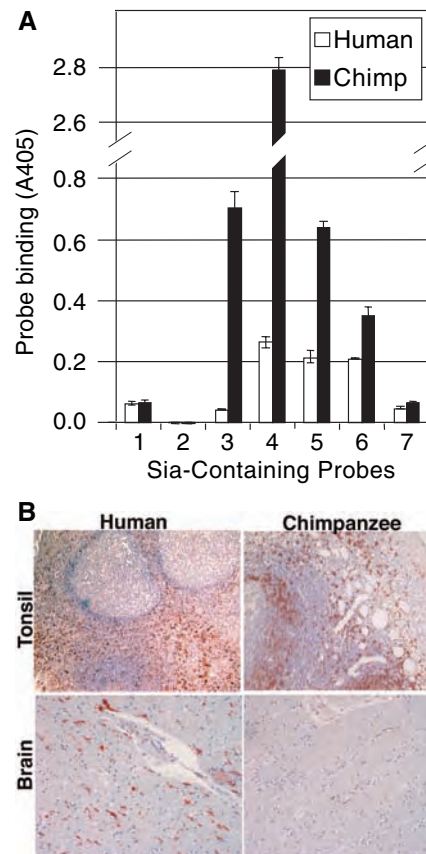
Fig. 2. Sia-recognition and expression patterns [4]. (**A**) Binding of recombinant human and chimpanzee Siglec-11 to Sia-containing probes: 1, Neu5Acα2-3Lac; 2, Neu5Acα2-6Lac; 3, Neu5Acα2−; 4, Neu5Gcα2−; 5, $(Neu5Ac\alpha2-8)_2$; 6, $(Neu5Ac\alpha2-8)_3$; and 7, $(Neu5Ac\alpha2-8)_{5-6}$. Error bars indicate standard deviation. A405, absorbance at 405 nm. (**B**) Examples of human and chimpanzee brain and tonsil tissue probed with antibody to human Siglec-11 (brown staining).

[1]Glycobiology Research and Training Center, [2]Department of Cellular and Molecular Medicine, [3]Department of Biological Sciences, [4]Department of Pathology, [5]Department of Medicine, University of California at San Diego, La Jolla, CA 92093, USA. [6]Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, MA 02141, USA.

*These authors contributed equally to this work.
†Present Address: Research Center for Glycoscience, National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan.
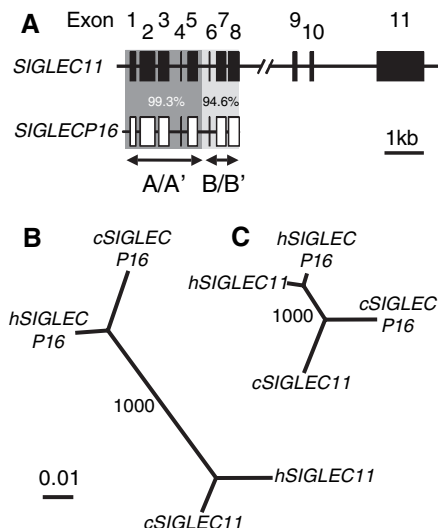‡To whom correspondence should be addressed. E-mail: a1varki@ucsd.edu

The mRNA of the *Arabidopsis* Gene *FT* Moves from Leaf to Shoot Apex and Induces Flowering

Tao Huang,[1] **Henrik Böhlenius,**[1] **Sven Eriksson,**[1]
François Parcy,[2] **Ove Nilsson**[1]*

Day length controls flowering time in many plants. The day-length signal is perceived in the leaf, but how this signal is transduced to the shoot apex, where floral initiation occurs, is not known. In *Arabidopsis*, the day-length response depends on the induction of the *FLOWERING LOCUS T* (*FT*) gene. We show here that local induction of *FT* in a single *Arabidopsis* leaf is sufficient to trigger flowering. The *FT* messenger RNA is transported to the shoot apex, where downstream genes are activated. These data suggest that the *FT* mRNA is an important component of the elusive "florigen" signal that moves from leaf to shoot apex.

In many plant species, the length of the day is a major environmental determinant controlling the time of flowering. For instance, *Ara-*

[1]Umeå Plant Science Centre, Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, S-90183, Umeå, Sweden. [2]Laboratoire de Physiologie Cellulaire Végétale, Département Réponse et Dynamique Cellulaires (DRDC/PCV), Unité Mixte de Recherche 5168 [(UMR) Joint Research Unit] Centre National de la Recherche Scientifique (CNRS), Commissariat à l'Energie Atomique (CEA), Institut National de la Recherche Agronomique (INRA), Université Joseph Fourier, 17 rue des Martyrs, bâtiment C2–38054, Grenoble Cedex 9, France.

*To whom correspondence should be addressed. E-mail: Ove.Nilsson@genfys.slu.se

bidopsis thaliana is a facultative long-day plant. Although it can flower in short days, it flowers much faster when days are long. The day-length signal is perceived in the leaf, where it induces a graft-transmissible signal that moves through the phloem sieve elements to the shoot apex, where flowering is initiated (*1*–*3*). This signal is sometimes referred to as the floral stimulus or florigen (*1*, *3*–*5*).

In *Arabidopsis*, two genes have been shown to be central for the photoperiodic response. Mutations in the genes *CONSTANS* (*CO*) and *FLOWERING LOCUS T* (*FT*) lead to late flowering under inductive long-day conditions, whereas flowering under noninductive short-

day conditions is only slightly affected (*6*). *CO* is expressed mainly in the leaf, where the CO protein is responsible for sensing the day-length signal (*7*–*10*). CO then induces *FT* in the leaf phloem (*8*, *9*). CO appears to act upstream of the graft-transmissible floral stimulus, but neither *CO* nor *FT* appears to be expressed in the shoot apex, where floral initiation occurs (*8*, *9*). CO can induce *FT* expression in the leaf phloem, but not when expressed from shoot apex–specific promoters (*9*, *11*). In contrast, *FT* expression in both the leaf and the shoot apex can trigger floral initiation (*9*). These data raise the possibility that the *FT* mRNA or the FT protein could be a part of the floral stimulus that moves from the leaf to the shoot apex.

Local induction of *FT*. To separate the role of *FT* from that of other genes that are induced in response to an increase in day length, we determined whether a local induction of *FT* in a single leaf is sufficient to induce flowering under noninductive short-day conditions. To test this, we constructed transgenic plants expressing *FT* and the reporter gene *GUS* under control of a heat shock–inducible promoter (*Hsp*) from soybean (*12*, *13*). We then heated to 37°C a single *Arabidopsis* leaf attached to a plant grown under short-day conditions (fig. S1); this heating induced a local activation of the *Hsp* promoter. We then monitored gene expression in the induced leaf and in microdissected shoot apices (*13*). Although heating the whole plant induced *GUS* transcription in the young leaf and the shoot
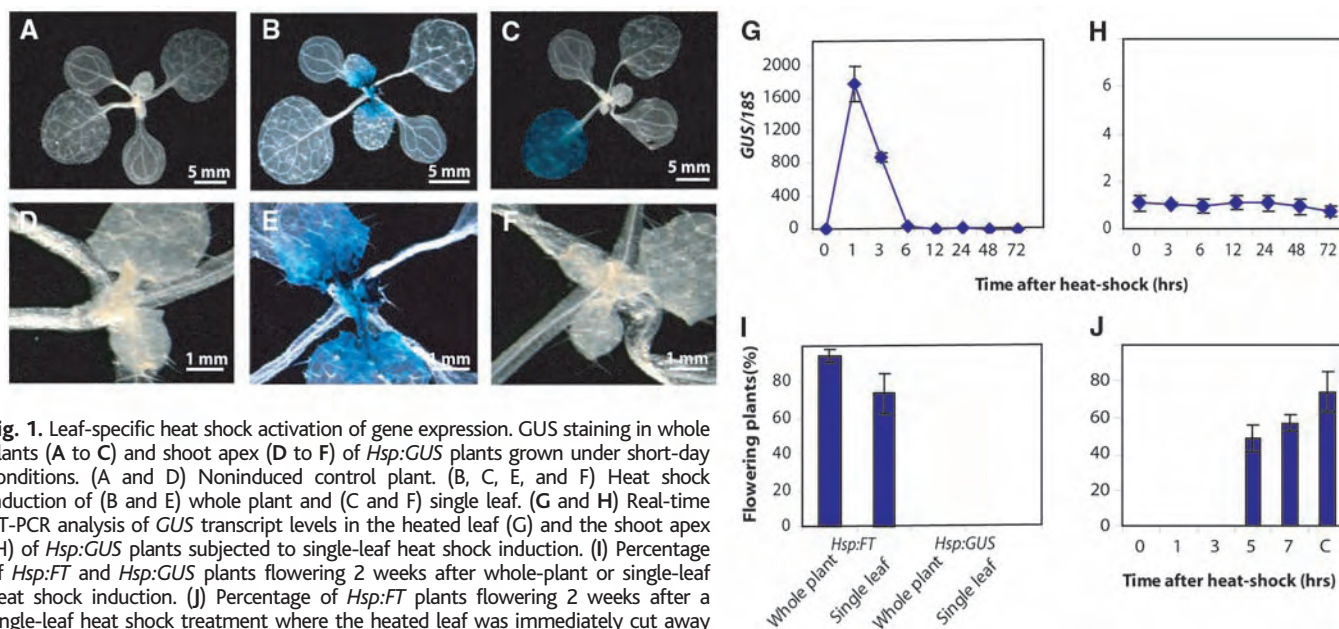
Fig. 1. Leaf-specific heat shock activation of gene expression. GUS staining in whole plants (**A** to **C**) and shoot apex (**D** to **F**) of *Hsp:GUS* plants grown under short-day conditions. (A and D) Noninduced control plant. (B, C, E, and F) Heat shock induction of (B and E) whole plant and (C and F) single leaf. (**G** and **H**) Real-time RT-PCR analysis of *GUS* transcript levels in the heated leaf (G) and the shoot apex (H) of *Hsp:GUS* plants subjected to single-leaf heat shock induction. (**I**) Percentage of *Hsp:FT* and *Hsp:GUS* plants flowering 2 weeks after whole-plant or single-leaf heat shock induction. (**J**) Percentage of *Hsp:FT* plants flowering 2 weeks after a single-leaf heat shock treatment where the heated leaf was immediately cut away (0 hour) or 1, 3, 5, and 7 hours later. Plants where the heated leaf was not cut away were used as positive controls [C in (J)]. For the *Hsp:GUS* control plants in (I) and (J), the first floral buds are visible after 7 weeks, at the same time as for wild-type plants grown in short days. Abscissa indicates chronological time after the start of the heat shock. Error bars, ±SD. Details of these experiments are described in (*13*).

apex (Fig. 1, A, B, D, and E), heating a single leaf induced strong transient *GUS* transcription in this leaf (Fig. 1, C and G), but no increase in *GUS* activity or mRNA level could be detected in the shoot apex (Fig. 1, F and H). This confirms that this heat shock system can be used to induce genes locally and that the *GUS* transcript appears incapable of moving from leaf to shoot apex.

We then used *Hsp:FT* transgenic plants to test whether a single pulse of *FT* transcription in leaves was sufficient to trigger flowering. Heating the whole plant caused 95% of the plants to develop visible flower buds 2 weeks after induction; control plants remained vegetative for another 3 weeks (Fig. 1I). However, the local heating of a single leaf was almost as efficient in inducing early flowering, leading to visible flower buds in 75% of the treated plants (Fig. 1I). We also tested whether this response was dependent on the presence of the endogenous *FT* gene. In the *ft* mutant background, there was no difference in the efficiency in which *Hsp:FT* induced flowering when the whole plant was heated, and only a slight reduction in efficiency when a single leaf was heated (fig. S2). Thus, a single pulse of *FT* induction in an individual leaf was sufficient to trigger flowering.

Movement of the *FT* mRNA. Because earlier studies suggested that *FT* might contribute to the mobile floral stimulus and that small RNA molecules can enter and move through the phloem of several plant species (*14–17*), we tested whether the *FT* transcript can travel.

In the *Hsp:FT Hsp:GUS* plants, the single-leaf heat shock treatment induced a strong transient expression of the transgenic *FT* transcript and the *GUS* transcript in the leaf (Fig. 2, A and C). An increase in the shoot apical levels of transgenic *FT* transcript could be detected 6 hours after the start of the leaf induction (Fig. 2B). This increase was not due to activation of the *Hsp* promoter by *FT*, because *Hsp:GUS* is not induced in the same transgenic plants (Fig. 2D). This suggests that the transgenic *FT* transcript, but not the *GUS* control transcript, can move from leaf to shoot apex. The heat shock treatment did not affect the expression of the endogenous *FT* gene in a transgenic *Hsp:GUS* plant (Fig. 2, E and F). However, in the *Hsp:FT Hsp:GUS* plants the levels of the endogenous *FT* transcript started to increase 6 to 12 hours after the leaf induction, both in leaves and in the shoot apex (Fig. 2, E and F). This suggests the existence of a positive autoregulatory loop where FT can induce, directly or indirectly, its own expression. Unlike transgenic *FT*, endogenous *FT* expression continues to increase in the induced leaf, even 3 days after induction, in spite of the fact that the plants are maintained in noninducing short-day conditions (Fig. 2E). This suggests that, once induced, *FT* can stably maintain its expression irrespective of day-length conditions. This finding explains classical experiments in which a leaf that has received a floral inductive signal stably continues to generate a graft-transmissible signal after up to seven successive graftings on multiple plants under noninductive condi-

tions (*1, 18*). Our finding of positive *FT* autoregulation also explains the phenomenon of indirect induction of flowering, that is, shoots induced to flower by grafting to donor shoots can themselves act as donors in subsequent grafts (*1*), which suggests that the floral stimulus can act in the leaves of these species to trigger its own synthesis. Our observations, therefore, give further support to the idea that *FT* is, at least partly, involved in the production of the classical floral stimulus.

The increased levels of endogenous *FT* transcript in the shoot apex could be due to transport of the leaf-induced *FT* and to de novo transcription in the apex. In order to distinguish between these two possibilities, we analyzed the activity of the *FT* promoter in heat shock–induced *Hsp:FT pFT:GUS* plants. *FT* promoter activity followed closely the levels of the endogenous transcript in the heat-shocked leaves (Fig. 2G). In the shoot apex, the *GUS* transcript levels increased between 6 and 12 hours (Fig. 2H), as did the levels of the endogenous *FT* transcript (Fig. 2F). Because the *GUS* transcript cannot move from leaf to shoot apex (Fig. 2, C and D), this result shows that a pulse of *FT* transcription in the leaves can induce *FT* promoter activity in the shoot apex. However, it does not exclude a contribution from transport of leaf-induced endogenous *FT* transcript.

To confirm movement of the *FT*-induced signal from the leaf to apex and to more closely analyze the kinetics of this movement, we removed the heat-shocked leaf at various
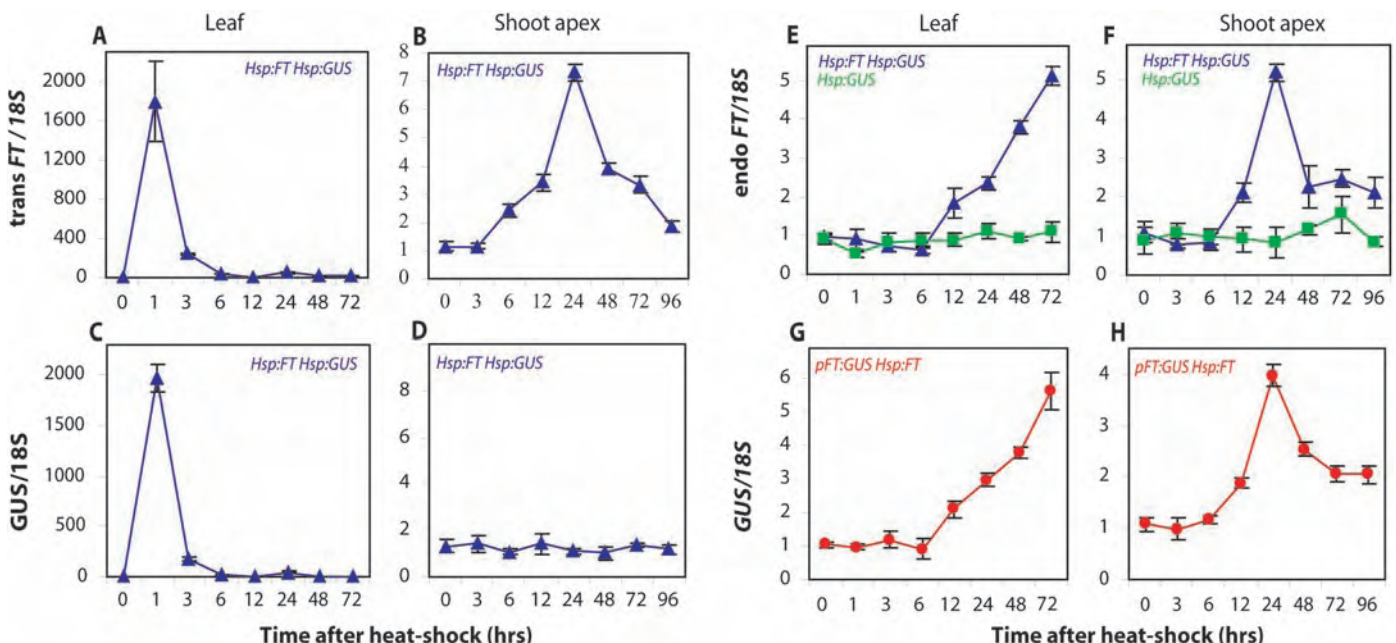
Fig. 2. *FT* transcript levels and *FT* promoter activity in heat shock–induced plants. *FT* and *GUS* transcript levels in the heated leaves (**A**, **C**, **E**, and **G**) and in the shoot apex (**B**, **D**, **F**, and **H**) of 16-day-old short day–grown plants where an individual leaf was subjected to heat shock induction. Chronological time after the start of the heat shock is shown on the *x* axis. (**A** and **B**) *FT* transcript levels derived from the *Hsp:FT* construct (trans *FT*). (**C** and **D**) *GUS* transcript levels derived form the *Hsp:GUS* construct. (**E** and **F**) *FT* transcript levels derived from the endogenous *FT* gene (endo *FT*) in *Hsp:FT Hsp:GUS* plants (blue triangles) and in *Hsp:GUS* control plants (green squares). (**G** and **H**) *GUS* transcript levels derived from the *pFT:GUS* construct. Error bars, ±SD. Details of these experiments are described in (*13*).
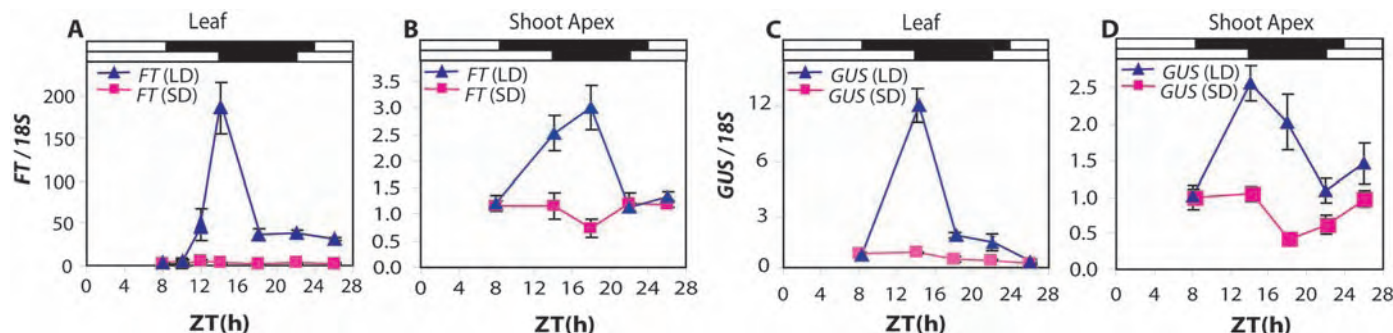
Fig. 3. *FT* transcript levels and *FT* promoter activity in response to a short-day to long-day shift. *pFT:GUS* plants were grown under short days for 4 weeks. At the end of the last short day, half of the plants were shifted to long days (blue triangles) giving a 5-hour extension of that day. The other half of the plants were kept in short days (red squares). The expression levels of the *FT* (**A** and **B**) and *GUS* (**C** and **D**) transcripts in leaf 5 or 6 (A and C) or the shoot apex (B and D) was determined by real-time RT-PCR analysis. Fifteen to 20 shoot apices or leaves were collected at the different time points, with three replicate samples. Error bars, ±SD. Black boxes indicate night and white boxes, light. ZT, zeitgeber time.

times after the heat shock induction. The induced leaf needed to be attached to the plant for more than 3 hours, but less than 5 hours, in order to induce an early flowering response (Fig. 1J). This correlates well with the first detectable occurrence of the transgenic *FT* transcript in the shoot apex 5 hours after heating (Fig. 2B). This also correlates well with previous estimates of the velocity of movement of the leaf derived floral stimulus [~2.4 to 3.5 mm/h (*3*)]. The heat-induced leaf blade is 6 to 7 mm from the shoot apex in our system, and the *FT* transcript has reached the apex 2 to 5 hours after maximum leaf induction (Fig. 2B); we calculate a velocity between 1.2 and 3.5 mm/h.

FT mRNA in the apex. To confirm that the levels of *FT* transcript are also increased in the apex of long day–induced "wild-type" plants, and to confirm the *FT* autoregulation, we analyzed the levels of *FT* and *GUS* transcript in a transgenic *pFT:GUS* plant after a short-day to long-day shift. At the end of the first long day, the levels of both the *FT* transcript and the *GUS* transcript had increased significantly in the leaf (Fig. 3, A and C). Likewise, both transcripts were increased in the shoot apex (Fig. 3, B and D), although at this resolution we could not determine whether the levels of the *FT* transcript increased before the *FT* promoter was activated in the shoot apex. Nevertheless, these results show that in normal long day–induced flowering, *FT* transcript levels increase in the shoot apex, and the *FT* promoter is induced. The most parsimonious explanation for these results is that this induction is at least partly caused by a transport of the *FT* transcript from leaf to shoot apex, followed by a positive autoregulation of the *FT* gene. Previous studies have failed to detect *FT* expression in the shoot apex (*8*, *9*); we attribute this discrepancy to the higher sensitivity of our reverse-transcription polymerase chain re-

action (RT-PCR) assay on microdissected shoot apices.

Downstream targets. To further characterize *FT*-induced floral induction, we looked at the activation of putative downstream targets. *APETALA 1* (*AP1*) expression is an early marker for reproductive development, already expressed at stage 1 of the induced floral primordium (*19*). The earliest signs of *AP1* induction at the apex could be seen 48 hours after *Hsp:FT* induction in the leaf (fig. S3A). This fits well with in situ hybridization data showing the first signs of *AP1* induction 48 to 72 hours after a short-day to long-day shift (*20*). *APETALA 3* (*AP3*) expression is induced at stage 3 of flower development (*21*). Consequently we could detect the first signs of *AP3* induction 72 hours after *FT* induction (fig. S3B). However, several genes showed a more rapid response to *FT* induction; *LEAFY* (*LFY*) (fig. S3C), *SUPPRESSOR OF OVER-EXPRESSION OF CONSTANS 1* (*SOC1*) (fig. S3D), *CAULIFLOWER* (*CAL*) (fig. S3E), and *FRUITFULL* (*FUL*) (fig. S3F) were all induced between 6 and 12 hours, suggesting that these genes are early targets in *FT*-induced floral induction. This is consistent with in situ hybridization data showing that an increase in *LFY* and *FUL* (*AGL8*) transcript levels can be detected 16 hours after a short-day to long-day shift (*20*).

Summary. Our data indicate that the *FT* mRNA is part of the mobile floral stimulus. However, we cannot exclude the possibility that the FT protein could also be moving and could be responsible for the floral induction. Neither can we exclude that FT may induce another gene or compound in the leaf that moves together with the *FT* transcript to induce flowering. In addition, the *FT* autoregulation demonstrated here could help to transduce the *FT* signal with a relay mechanism involving subsequent *FT* reinductions. Nevertheless, the simplest explanation of our data is that the *FT* mRNA constitutes an important part of the floral

stimulus that moves from leaf to shoot apex.

Note added in proof: It has now been shown that FT can act in the shoot apex by controlling the activity of the shoot apex–expressed transcription factor FD (*22*, *23*).

References and Notes

1. J. A. D. Zeevaart, *Annu. Rev. Plant Physiol.* **27**, 321 (1976).
2. J. E. Knott, *Proc. Am. Soc. Hort. Sci.* **31**, 152 (1934).
3. M. Chailakhyan, *Dokl. Acad. Sci. URSS* **13**, 77 (1936).
4. L. Corbesier, G. Coupland, *Plant Cell Environ.* **28**, 54 (2005).
5. J. Colasanti, V. Sundaresan, *Trends Biochem. Sci.* **25**, 236 (2000).
6. M. Koornneef, C. J. Hanhart, J. H. van der Veen, *Mol. Gen. Genet.* **229**, 57 (1991).
7. M. J. Yanovsky, S. A. Kay, *Nat. Rev. Mol. Cell Biol.* **4**, 265 (2003).
8. S. Takada, K. Goto, *Plant Cell* **15**, 2856 (2003).
9. H. An *et al.*, *Development* **131**, 3615 (2004).
10. F. Valverde *et al.*, *Science* **303**, 1003 (2004).
11. B. G. Ayre, R. Turgeon, *Plant Physiol.* **135**, 2271 (2004).
12. K. Severin, F. Schoffl, *Plant Mol. Biol.* **15**, 827 (1990).
13. Material and methods are available as supporting material on *Science* Online.
14. B.-C. Yoo *et al.*, *Plant Cell* **16**, 1979 (2004).
15. M. Kim, W. Canio, S. Kessler, N. Sinha, *Science* **293**, 287 (2001).
16. B. Xoconostle-Cázares *et al.*, *Science* **283**, 94 (1999).
17. B. Ding, A. Itaya, Y. Qi, *Curr. Opin. Plant Biol.* **6**, 596 (2003).
18. J. A. D. Zeevaart, in *Handbook of Flowering*, A. H. Halevy, Ed. (CRC Press, Boca Raton, FL, 1985), pp. 239–252.
19. M. A. Mandel, C. Gustafson-Brown, B. Savidge, M. F. Yanofsky, *Nature* **360**, 273 (1992).
20. F. D. Hempel *et al.*, *Development* **124**, 3845 (1997).
21. T. Jack, L. L. Brockman, E. M. Meyerowitz, *Cell* **68**, 683 (1992).
22. M. Abe *et al.*, *Science* **309**, 1052 (2005).
23. P. A. Wigge *et al.*, *Science* **309**, 1056 (2005).
24. We thank K. Goto for the gift of the *pFT:GUS* seeds and E. Ögren and K. Bergman for technical assistance. This work was supported by an INGVAR grant (Individual Grants for the Advancement of Research Leaders) from the Swedish Foundation for Strategic Research to O.N.

Spin Disorder on a Triangular Lattice

**Satoru Nakatsuji,[1]* Yusuke Nambu,[1] Hiroshi Tonomura,[1]
Osamu Sakai,[1] Seth Jonas,[3] Collin Broholm,[3,4]
Hirokazu Tsunetsugu,[2] Yiming Qiu,[4,5] Yoshiteru Maeno[1,6]**

As liquids crystallize into solids on cooling, spins in magnets generally form periodic order. However, three decades ago, it was theoretically proposed that spins on a triangular lattice form a liquidlike disordered state at low temperatures. Whether or not a spin liquid is stabilized by geometrical frustration has remained an active point of inquiry ever since. Our thermodynamic and neutron measurements on $NiGa_2S_4$, a rare example of a two-dimensional triangular lattice antiferromagnet, demonstrate that geometrical frustration stabilizes a low-temperature spin-disordered state with coherence beyond the two-spin correlation length. Spin liquid formation may be an origin of such behavior.

Electronic magnetic moments (spins) in magnets generally develop periodic order at low temperatures. When such order is suppressed, however, qualitatively new quantum phases can emerge. For example, quantum spin liquids appear in quasi–one-dimensional spin chains when fluctuations enhanced by low dimensionality destabilize static correlations (1). In higher dimensions, magnetic order may also be suppressed because of the geometry of the crystal lattice. Such an effect occurs, for example, in an antiferromagnet with a structure formed by arrays of triangles. Antiferromagnetically interacting spins on the vertices of triangles cannot simultaneously satisfy all pairwise interactions and may remain disordered well below the conventional ordering scale set by the Weiss temperature, θ_W (2). Thus, "geometrical frustration" promotes high degeneracy among competing low-temperature phases, enhances quantum fluctuations, and may lead to unconventional quantum phenomena.

In two dimensions, the simplest form of a geometrically frustrated lattice is a triangular lattice with a single magnetic atom per unit cell. This is the system for which a quantum spin-disordered state in more than one dimension was first proposed more than three decades ago (3). Since then, extensive theoretical and experimental research has been carried out in pursuit of novel ground states without magnetic order in

two dimensions. Although it is now believed that triangular lattice antiferromagnets with nearest-neighbor coupling exhibit 120° spin order (4–6), recent theories suggest that interactions beyond nearest-neighbor exchange, such as longer range and multiple-spin exchange interactions, may lead to a quantum spin-disordered ground state (7, 8). Experimentally, however, only a few candidates for spin-disordered states in two dimensions have been reported, in an organic material with a distorted triangular lattice (9), in Kagomé-related antiferromagnets (10, 11), and in a low-density solid ³He film absorbed on a graphite surface (12). So far unanswered is the question of whether an insulating bulk solid with an exact triangular lattice

of well-defined localized moments can exhibit a novel ground state without magnetic order.

We demonstrate that high-quality samples of $NiGa_2S_4$, a bulk-insulating antiferromagnet on an exact triangular lattice, exhibit a spin-disordered state in two dimensions. Despite strong antiferromagnetic (AF) interactions, no magnetic long-range order was observed down to 0.35 K, where we instead found nanoscale quasi-static correlation. The spin-disordered state appeared on cooling through highly degenerate states with an entropy plateau and exhibited gapless, linearly dispersive modes, suggesting coherence beyond the two-spin correlation length. The gapless excitations were insensitive to a magnetic field but sensitive to impurities. These observations indicate the absence of canonical spin glass freezing in the bulk, raising the possibility of a spin liquid state with slow dynamics.

$NiGa_2S_4$ is a layered chalcogenide magnetic insulator with a stacked triangular lattice of Ni spins (13) (Fig. 1A). The structure is highly two-dimensional (2D) and the central NiS_2 layer is isostructural with the CoO_2 layer in superconducting $Na_xCoO_2 \cdot yH_2O$ (14). Magnetism is associated with Ni^{2+} with the electronic configuration $t_{2g}^6 e_g^2$ and spin $S = 1$. The unit cell contains a single Ni^{2+} atom, so the intra- and inter-plane Ni separations are simply given by the lattice parameters $a = 0.3624$ nm and $c = 1.1999$ nm.

In Fig. 2A, the temperature dependence of the magnetic susceptibility, $\chi(T) \equiv M(T)/B$, and its inverse, $\chi^{-1}(T)$, where T is the temperature, are presented. The applied field B

[1]Department of Physics, [2]Yukawa Institute for Theoretical Physics, Kyoto University, Kyoto 606-8502, Japan. [3]Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD 21218, USA. [4]NIST Center for Neutron Research, National Institute of Standards and Technology (NIST), Gaithersburg, MD 20899, USA. [5]Department of Materials Science and Engineering, University of Maryland, College Park, MD 20742, USA. [6]International Innovation Center, Kyoto University, Kyoto 606-8501, Japan.

*To whom correspondence should be addressed. E-mail: nakatsuji@scphys.kyoto-u.ac.jp

Fig. 1. Crystal and spin structures of $NiGa_2S_4$. (**A**) The slab structure of $NiGa_2S_4$ consists of a central NiS_2 layer of edge-sharing NiS_6 octahedra (red) and top and bottom sheets of GaS_4 tetrahedra (green). The slabs are stacked along the c axis and separated from each other by a van der Waals gap. Polycrystalline samples of $NiGa_2S_4$ were synthesized by heating the mixture of elements Ni, Ga, and S in evacuated silica ampoules at 900°C. Our powder x-ray measurements at room temperature and neutron diffraction measurements in the temperature range 1.5 K to 300 K confirmed that $NiGa_2S_4$ retains the trigonal crystal structure down to 1.5 K with $P\bar{3}m1$ symmetry. The refinements are consistent with the room-temperature structure (13). (**B**) The short-range correlated spin structure on the triangular Ni lattice, including the incommensurability observed by magnetic neutron diffraction. It approximates four independent 120° structures on the colored sublattices with lattice parameters $2a$. The in-plane correlation length is 2.5(3) nm, and the correlation time exceeds 0.3 ns. There are weak ferromagnetic correlations between nearest-neighbor planes.

was 7 T, a field at which the magnetization M remains proportional to B. No difference was found between field-cooled and zero field-cooled data. Susceptibility above 150 K followed the Curie-Weiss law: $\chi^{-1}(T) = (T - \theta_W)/C$, and the Curie constant C corresponds to an effective moment of 2.81(3) Bohr magnetons (μ_B), consistent with spin-1 ($S = 1$) Ni^{2+}. The Weiss temperature, $\theta_W = -80(2)$ K, indicates strong AF interactions. However, in a field of 7 T, no sharp magnetic anomaly or field hysteresis was observed down to 1.8 K. Instead, the susceptibility smoothly increased on cooling and passed through a shallow and broad maximum at $T \sim 10$ K before approaching a finite low-T limit. These data exclude the possibility of a conventional spin-gap system in which the susceptibility vanishes exponentially at low temperatures.

Specific heat data, $C_p(T)$, also showed no evidence of a phase transition for fields between 0 T and 7 T and temperatures from 175 K down to 0.35 K (Fig. 2B, inset). These data indicate a disordered low-temperature state without conventional AF order. We probed the density of states for spin excitations through the magnetic specific heat $C_M(T)$ after subtraction of the lattice contribution (15). $C_M(T)$ (Fig. 2B) exhibited an unusual double-peak structure. One peak was broad and centered around $T = |\theta_W|$, below which the susceptibility was suppressed compared to the one expected from the high-temperature Curie-Weiss law (Fig. 2A). The lower temperature peak featured a prominent rounded maximum at $T_{peak} \approx 10$ K, where $\chi(T)$ also showed a broad maximum. The entropy S_M, obtained through integration of C_M/T (Fig. 2B), correspondingly had a plateau at $S_M \approx \frac{1}{3}R\ln3$ before high-T saturation at $S_M \approx R\ln3$, where R is the gas constant. This indicates high degeneracy of low-energy states because of magnetic frustration and is similar to predictions for the spin-1/2 antiferromagnet on the Kagomé lattice of corner-sharing triangles (16).

The specific heat exhibited power-law behavior at low temperatures (Fig. 3). The data between 0.35 and 4.0 K are well fitted by a power law $C_M = AT^\alpha$, where A is a constant and α is 2.001(5) (Fig. 3A). Quadratic temperature dependence through one decade indicates the presence of gapless and linearly dispersive modes in two dimensions. A gapless spectrum is furthermore consistent with a finite value of the susceptibility in the low-T limit. If the low-T specific heat peak were associated with AF ordering of individual $S = 1$ spins, it would be suppressed under a magnetic field of $k_B T_{peak}/g\mu_B \sim 7$ T, where k_B is the Boltzmann constant and g is the g factor. Instead, the specific heat was unaffected by fields up to 7 T, indicating that the low-T peak reflects short-range correlations among composite degrees of freedom that do not directly couple to a uniform field (17). It follows that the linearly dispersive

excitations implied by $C_M \propto T^2$ are collective modes of moment-free spin clusters. Likewise, the specific heat peak at high $T \sim |\theta_W|$ and the concomitant suppression of the susceptibility compared to the one expected from high-temperature Curie-Weiss law are attributable to the formation of incoherent moment-free spin clusters. Similar thermodynamic properties have been reported for the Kagomé-related antiferromagnets, $SrCr_{9p}Ga_{12-9p}O_{19}$ (10, 17), and for deuteronium jarosite (11).

Although no susceptibility anomaly was observed at 7 T, weak field dependence and hysteresis was observed at low fields for temperatures below a freezing temperature $T_f = 8.5$ K (Fig. 3B). Previous work indicates that minute levels of quenched disorder can induce spin freezing in geometrically frustrated magnets (2). However, the bifurcation of $\chi(T)$ in $NiGa_2S_4$ is distinct from previous observations in frustrated magnets in that it involves only a small component of the whole susceptibility
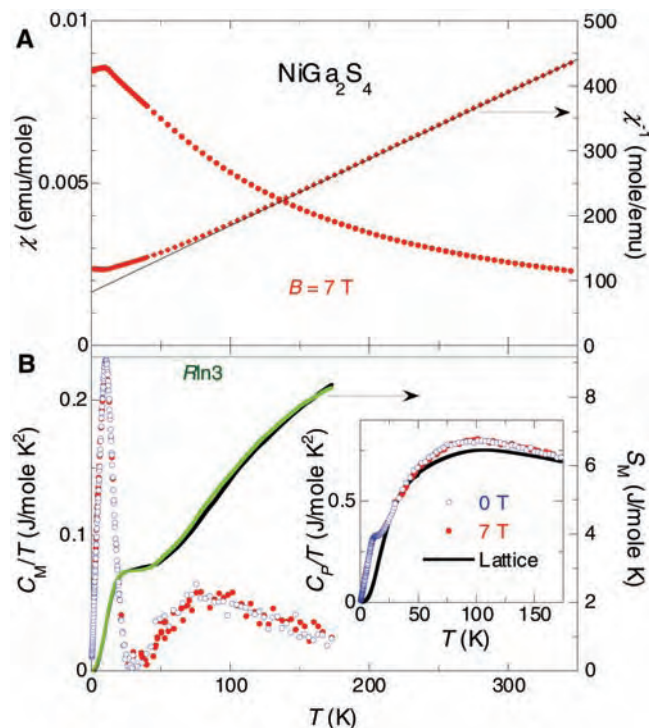
Fig. 2. Temperature dependence of the susceptibility and specific heat of polycrystalline $NiGa_2S_4$. (**A**) The susceptibility χ (circles, left axis) and inverse susceptibility χ^{-1} (diamonds, right axis), measured at 7 T with a superconducting quantum interference device (SQUID) magnetometer (Quantum Design, model MPMS-XL). The solid line is the Curie-Weiss fit. emu, electromagnetic unit. (**B**) The magnetic part of the specific heat divided by temperature C_M/T (left axis) at zero field (blue open circles) and at 7 T (red solid circles) and the entropy S_M (right axis) at 0 T (green line) and at 7 T (black line). The horizontal line indicates $S_M = R\ln3$. Inset: C_p/T for $NiGa_2S_4$ at 0 T (blue open circles) and 7 T (red solid circles) and its lattice part C_L/T [black solid line (15)].
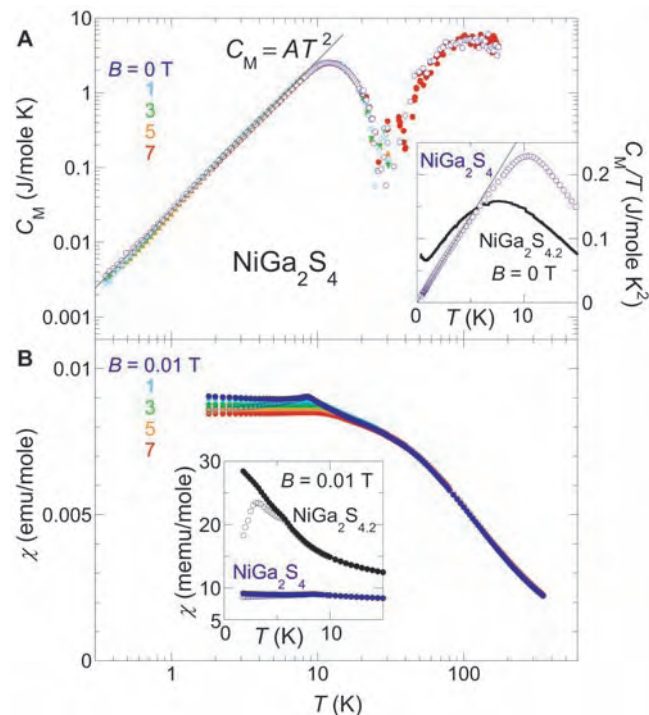
Fig. 3. Magnetic field dependence of the specific heat and susceptibility of $NiGa_2S_4$. (**A**) The magnetic part of the specific heat C_M under different fields versus temperature in full logarithmic scale. Inset: C_M/T versus temperature under zero field for $NiGa_2S_4$ and the sample with nominal doping of excess 5% sulfur, $NiGa_2S_{4.2}$. The solid lines show the T^2 dependence of C_M. (**B**) The susceptibility χ under different fields. Inset: The low-temperature part of χ for $NiGa_2S_4$ and $NiGa_2S_{4.2}$ at 0.01 T. Both field-cooled (solid circles) and zero field-cooled (open circles) data are shown.

(∼5%). The following observations indicate that it may in fact be a surface effect. First, intentional impurity doping by 5% excess sulfur leads to a substantial enhancement of freezing behavior (18). A large value of $C_M(T)/T$ (∼70 mJ/mol K²) was induced for $T \to 0$ and hysteresis in $\chi(T)$ was strongly enhanced (Fig. 3, inset). Second, when a field around the critical value $B_c \sim 7$ T was applied, the hysteretic part of the susceptibility completely disappeared, whereas the specific heat, $C_M(T)$, showed only negligible field dependence in the same B-T range. Because magnetism is extremely sensitive to imperfection, the different field effects of the two thermodynamic quantities, together with $C_M(T)/T \to 0.0(1)$ mJ/mol K² as $T \to 0$ K, indicate that the hysteretic part of the susceptibility is associated with a very small density of impurities. In addition, the value of B_c implies that impurity spins are individual $S = 1$ objects, as this value agrees with the magnetic field where $S = 1$ impurity spins become polarized, $B_c \sim k_B T_f / g\mu_B$. Third, the density of spins involved in freezing can be estimated from the difference between $\chi(T)$ for $B = 0$ T and 7 T, which follows a Curie-Weiss law with $\theta_W = 9$ K and an effective moment corresponding to 300 parts per million (ppm) of the Ni sites. This concentration matches that of surface spins in the ∼10 μm³ grains of our sample. Finally, this small density of impurities is consistent with diffraction measurements. No trace of disorder or impurities was found by x-ray or neutron powder diffraction, and Rietveld refinement of neutron powder diffraction data is consistent with the nominal stoichiometry.

To investigate low-energy spin correlations, we performed a magnetic neutron scattering experiment (19). Figure 4 shows the difference in elastic powder neutron scattering between $T =$ 1.5 K and 15 K. The signal was resolution-limited in energy [$\delta E = 0.10$ meV, where δE is full width at half maximum (FWHM) energy resolution], which allows us to conclude that the

correlations persist on a time scale that exceeds 0.3 ns. The absence of a difference signal in the $Q \to 0$ limit, where Q is the scattering wave vector, indicates a correlated state built from moment-free spin clusters. Classical spins on a triangular lattice with nearest-neighbor AF interactions should develop correlations of the 120° variety with wave vector $q = (\frac{1}{3}, \frac{1}{3}, 0)$. NiGa₂S₄ exhibited a peak at $Q_M =$ 0.57(1) Å⁻¹ ≈ $2\pi/3a = 0.5779$ Å⁻¹, which is the length of $q = (\frac{1}{6}, \frac{1}{6}, 0)$. To understand this result, we compared the data to the spherical average of magnetic scattering from a quasi-2D magnet:

$$\frac{\partial \sigma}{\partial \Omega} = r_0^2 \left| \frac{g}{2} F(Q) \right|^2 N \sum_\tau (|m_q|^2 - |\hat{Q} \cdot m_q|^2)$$

$$\times [1 + 2\alpha \cos(Q \cdot c)] \frac{A^* \kappa^2 / \pi}{[(Q - \tau \pm q)^2 + \kappa^2]^2}$$

$$(1)$$

Here σ is the scattering cross section, Ω is the solid angle of detection, $r_0 = 5.4 \times 10^{-15}$ m, F is the magnetic form factor for Ni²⁺, N is the number of magnetic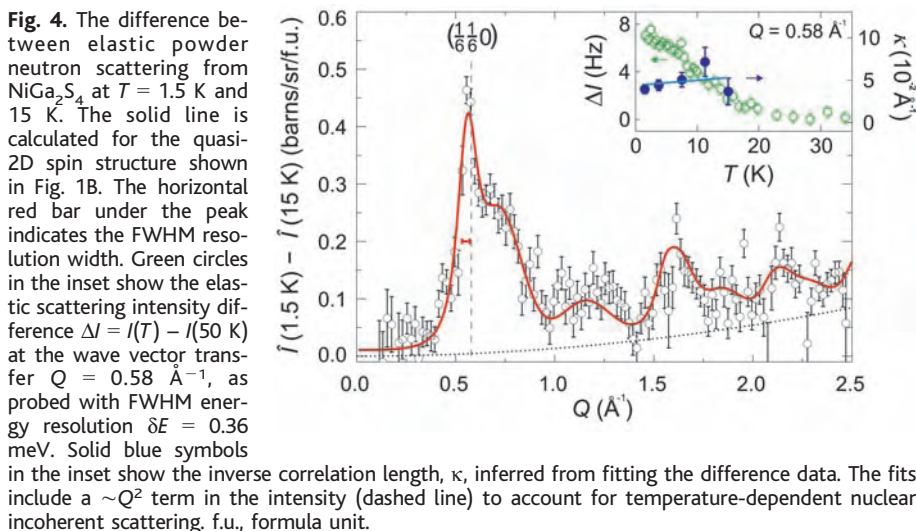 ions, A^* is the area of the Brillouin zone, and the summation is over 2D reciprocal lattice vectors τ. The spin expectation value on site, r, averaged over a time interval that exceeds 0.3 ns, is given by $\langle S_r \rangle = m_q e^{iq \cdot r} + m_q^* e^{-iq \cdot r}$, where m_q is a complex vector. In Eq. 1, κ is the inverse 2D correlation length and $\alpha = \langle S_0 \cdot S_{\pm c} \rangle \langle S_0 \cdot S_0 \rangle^{-1}$ represents nearest-neighbor inter-plane correlations. When q was not $(\frac{1}{3}, \frac{1}{3}, 0)$, there were three symmetry-related wave vector domains. The line in Fig. 4 corresponds to $q = (\eta, \eta, 0)$ with $\eta = 0.158(1)$ and $m_q = \hat{x} m_{qx} + i\hat{y}m_{qy} + \hat{z}m_{qz}$, where $m_{qx} = 0.31(3)$, $m_{qy} = 0.43(4)$, and $m_{qz} = 0.0(1)$. Here $\hat{z} \parallel c$ and $\hat{x}$, $\hat{y}$ are parallel and perpendicular to q, respectively. The corresponding time- and site-averaged spin is $|\langle S \rangle| = \sqrt{(1/N) \sum_r \langle S_r^2 \rangle} = \sqrt{2|m_q|} =$ 0.75(8). This value is notably reduced from $S = 1$ for Ni²⁺ because of quantum fluctuations

enhanced by frustration and low dimensionality. It is also important that the in-plane correlation length, $\xi = \kappa^{-1} = 2.5(3)$ nm, corresponds to 6.9(8) triangular lattice spacings and ∼2.5 × 10⁻⁴ times the powder sample grain size. Ferromagnetic inter-plane correlations were limited to nearest neighbors as indicated by $\alpha = 0.25(5)$.

The incommensurate short-range order inferred from the diffraction data (Fig. 1B) has various possible explanations. If third-neighbor AF interactions, J_3, dominate, magnetic orders with $q = (\frac{1}{6}, \frac{1}{6}, 0)$ and/or $(\frac{1}{3}, \frac{1}{3}, 0)$ will be degenerate at the mean field level. Nearest neighbor couplings, J_1, lift the degeneracy. When J_1 is ferromagnetic, a $(\frac{1}{6}, \frac{1}{6}, 0)$ order is favored, but the wave vector will be shifted to an incommensurate value (η, η, 0), as in our diffraction data. From the experimental value of $\eta = 0.158(1)$, mean field theory yields $J_1/J_3 =$ −0.2(1). Indeed, because the nearest-neighbor Ni-S-Ni bond angle was near 90°, J_1 should be small and potentially ferromagnetic. Alternatively, proximity of the NiS₂ layer to a metal-insulator transition may produce multiple-spin exchange that stabilizes a larger unit cell.

Although the spin correlation length, $\xi =$ 2.5(3) nm, is rather short, specific heat data indicate a much longer coherence length for low-energy excitations. When coherent propagation of excitations is limited up to a length scale L_0 at $T = 0$, the specific heat deviates from the low-T asymptotic form as $C_M/R =$ $-(\sqrt{3}\pi/2)(a/L_0)^2 + [3\sqrt{3}\zeta(3)/2\pi](ak_B T/\hbar D)^2$ at $hD/(L_0 k_B) \ll T \ll \theta_W$, where D is the spin stiffness constant (20). Least-squares fitting of this expression to the specific heat data in Fig. 3 yields a lower bound, $L_0 > 200$ nm. Clearly, this is far greater than the two-spin correlation length, $\xi = 2.5(3)$ nm, determined by neutron diffraction, and it indicates that coherent low-energy modes are a bulk effect in NiGa₂S₄.

These experimental results demonstrate that NiGa₂S₄ has unique low-temperature properties: (i) absence of conventional magnetic order, which is replaced by incommensurate short-range order with nanoscale correlation; (ii) absence of canonical bulk spin glass freezing; (iii) gapless coherent excitations in two dimensions that are insensitive to field but sensitive to impurities; and (iv) highly degenerate low-energy states as indicated by an entropy plateau. These observations place strong constraints on possible ground states. One likely candidate is a spin liquid with no conventional magnetic long-range order as characterized by resolution-limited Bragg peaks, no static spin freezing, and an absence of spin dimerization (i.e., valence bond solid order) that would induce a finite spin gap. Observations (i), (ii), and (iii) are consistent with these criteria. Moreover, observations (iii) and (iv) are consistent with the type of spin liquid that is expected to result from quantum effects in a highly degenerate manifold induced by mag-

Fig. 4. The difference between elastic powder neutron scattering from NiGa₂S₄ at $T = 1.5$ K and 15 K. The solid line is calculated for the quasi-2D spin structure shown in Fig. 1B. The horizontal red bar under the peak indicates the FWHM resolution width. Green circles in the inset show the elastic scattering intensity difference $\Delta I = I(T) - I(50$ K) at the wave vector transfer $Q = 0.58$ Å⁻¹, as probed with FWHM energy resolution $\delta E = 0.36$ meV. Solid blue symbols in the inset show the inverse correlation length, κ, inferred from fitting the difference data. The fits include a ∼Q^2 term in the intensity (dashed line) to account for temperature-dependent nuclear incoherent scattering. f.u., formula unit.

netic frustration. For example, recent theoretical work on the single-band Hubbard model on the triangular lattice suggests that a gapless spin liquid phase is realized at intermediate correlations (8). It is also predicted that the spin liquid phase has both spin-zero and non-zero low energy excitations (21) and incommensurate short-range correlation (22), which may be related to these observations.

Another possibility is a delicate form of long-range order that is dynamically inaccessible in real materials because of impurity pinning close to quantum criticality. Other possible ground states include a Kosterlitz-Thouless (KT) phase driven by two-valued vortices with the transition around T_{peak} (23) and a long-range order without long-range spin correlations such as in a spin nematic (24). The KT phase discussed here is not a conventional one but has a finite spin-correlation length, consistent with our observations. The neutron scattering data show that a large part of the magnetic spectral weight is associated with a long time scale that exceeds 0.3 ns. Experiments characterizing slow spin dynamics are in progress to distinguish between these exotic potential ground states in NiGa$_2$S$_4$.

Like the spin-1/2 chain in one dimension (1), the triangular lattice antiferromagnet, which is the only geometrically frustrated 2D Bravais lattice, plays a central role in the search for cooperative phenomena in two dimensions. With AF nearest-neighbor interactions, Ising spins have residual entropy (25), and classical Heisenberg spins have a KT-type transition and noncolinear Néel order at $T = 0$ (23). In the extreme quantum spin-1/2 case, a superposition of singlet coverings constituting a resonating valence bond state was first introduced to describe the ground state of the triangular lattice antiferromagnet (3). Although it is now believed that a nearest-neighbor Heisenberg spin-1/2 model has 120° order at $T = 0$ (4–6), one may expect various exotic phases in real materials, driven by other effects such as longer range interactions, multiple-spin exchange, and structural dimerization. Consequentially, there have been intensive experimental searches for an ideal quasi-2D triangular lattice antiferromagnet with a small spin quantum number ($S \leq 1$) (26). Several spin-1/2 triangular lattices have been reported to show nonordered gapless phases, but all the materials so far have either a distorted structure (9, 27) or are only available in minute quantities, which limits accessibility for many experimental techniques (12).

To our knowledge, NiGa$_2$S$_4$ is the first low spin ($S \leq 1$), quasi-2D bulk magnet on an exact regular triangular lattice. Our measurements on NiGa$_2$S$_4$ demonstrate that geometrical frustration can be used to stabilize a spin-disordered state at low temperatures. A strongly fluctuating spin state has been discovered in a triangular lattice plane that is isostructural to the superconducting CoO$_2$ layer of Na$_x$CoO$_2$·yH$_2$O

(14) and isostoichiometric with NiS$_2$, a metal-insulator transition system (28). It may thus be possible to drive NiGa$_2$S$_4$ into a conducting state in which the magnetic fluctuations have a qualitative effect on carrier correlations.

References and Notes

1. I. Affleck, *J. Phys. Condens. Matter* **1**, 3047 (1989).
2. A. P. Ramirez, in *Handbook of Magnetic Materials*, K. J. H. Buschow, Ed. (Elsevier Science, Amsterdam, 2001), vol. 13, pp. 423–520.
3. P. W. Anderson, *Mater. Res. Bull.* **8**, 153 (1973).
4. D. A. Huse, V. Elser, *Phys. Rev. Lett.* **60**, 2531 (1988).
5. B. Bernu, C. Lhuillier, L. Pierre, *Phys. Rev. Lett.* **69**, 2590 (1992).
6. L. Capriotti, A. E. Trumper, S. Sorella, *Phys. Rev. Lett.* **82**, 3899 (1999).
7. G. Misguich, C. Lhuillier, in *Frustrated Spin Systems*, H. T. Diep, Ed. (World-Scientific, Singapore, 2004), pp. 229–306.
8. H. Morita, S. Watanabe, M. Imada, *J. Phys. Soc. Jpn.* **71**, 2109 (2002).
9. Y. Shimizu, K. Miyagawa, K. Kanoda, M. Maesato, G. Saito, *Phys. Rev. Lett.* **91**, 107001 (2003).
10. A. P. Ramirez, G. P. Espinosa, A. S. Cooper, *Phys. Rev. Lett.* **64**, 2070 (1990).
11. A. S. Wills, A. Harrison, S. A. M. Mentink, T. E. Mason, Z. Tun, *Europhys. Lett.* **42**, 325 (1998).
12. K. Ishida, M. Morishita, K. Yawata, H. Fukuyama, *Phys. Rev. Lett.* **79**, 3451 (1997).
13. H. D. Lutz, W. Buchmeier, H. Siwert, *Z. Anorg. Allg. Chem.* **533**, 118 (1986).
14. K. Takada *et al.*, *Nature* **422**, 53 (2003).
15. The specific heat, C_P, was measured by a thermal relaxation method down to 0.35 K under fields up to 7 T. In order to estimate the lattice part of the specific heat, C_L, we measured C_P for the isostructural nonmagnetic analog ZnIn$_2$S$_4$ and obtained the thermal variation of the Debye temperature $\theta_D(T)$ using the Debye equation (29). $\theta_D(T)$ of NiGa$_2$S$_4$ was then estimated by applying a scale factor according to $\theta_D \propto M_0^{-1/2}V_0^{-1/3}$, where M_0 and V_0 are molar mass and volume, respectively. Finally, C_L was estimated by converting the scaled $\theta_D(T)$ data into specific heat. The magnetic part, C_M, was estimated as the difference between the total specific heat, C_P, and the lattice estimation C_L.
16. P. Sindzingre *et al.*, *Phys. Rev. Lett.* **84**, 2953 (2000).
17. A. P. Ramirez, B. Hessen, M. Winklemann, *Phys. Rev. Lett.* **84**, 2957 (2000).
18. J. Mydosh, *Spin Glasses: An Experimental Introduction* (Taylor and Francis, London, 1993).
19. Neutron diffraction experiments were performed on the Disc Chopper Spectrometer at NIST with 3.55-meV neutrons and an elastic energy resolution of 0.10 meV. Measurements versus T were carried out on the cold neutron triple axis spectrometer SPINS, with an energy of 5 meV and an energy resolution of 0.36 meV.
20. High degeneracy of low-temperature states is indicated not only by the large value of the entropy plateau at $S_M \approx 1/3 R\ln3$ at $T \sim 20$ K but also by the stiffness constant D. For ordinary antiferromagnets that order at $T \sim |\theta_W|$, the stiffness D_0 is estimated by the relation $D_0^2 \approx [3\sqrt{3}\zeta(3)/4\pi](ak_B\theta_W/\hbar)^2/\ln(2S + 1)$, with $\zeta(3) = 1.202$. In our case, the observed D was 850 m/s, nearly three times smaller than the expected D_0 of 2500 m/s, indicating softening due to magnetic frustration.
21. M. Imada, T. Mizusaki, S. Watanabe, available at http://arxiv.org/abs/cond-mat/0307022 (2003).
22. T. Kashima, M. Imada, *J. Phys. Soc. Jpn.* **70**, 3052 (2001).
23. H. Kawamura, S. Miyashita, *J. Phys. Soc. Jpn.* **53**, 4138 (1984).
24. P. Chandra, P. Coleman, *Phys. Rev. Lett.* **66**, 100 (1991).
25. G. H. Wannier, *Phys. Rev.* **79**, 357 (1950).
26. M. F. Collins, O. A. Petrenko, *Can. J. Phys.* **75**, 605 (1997).
27. K. Takeda, K. Miyake, K. Takeda, K. Hirakawa, *J. Phys. Soc. Jpn.* **61**, 2156 (1992).
28. J. A. Wilson, in *Metallic and Nonmetallic States of Matter*, P. P. Edwards, C. N. R. Rao, Eds. (Taylor and Francis, London, 1985), pp. 215–260.
29. J. A. Beattie, *J. Math. Phys.* **6**, 1 (1926/1927).
30. We thank J. Y. Chan, S. Fujimoto, K. Ishida, K. Kitagawa, R. T. Macaluso, D. E. MacLaughlin, R. Moessner, S. Sondhi, and O. Tchernyshyov for fruitful discussions. Work at Kyoto University supported in part by Grants-in-Aid for Scientific Research from the Japan Society for the Promotion of Science and for the 21st Century Center of Excellence "Center for Diversity and Universality in Physics" from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan and by the Inamori Foundation. Work at Johns Hopkins University was supported by the U.S. Department of Energy.

Conversion of Zinc Oxide Nanobelts into Superlattice-Structured Nanohelices

Pu Xian Gao,[1] Yong Ding,[1] Wenjie Mai,[1] William L. Hughes,[1] Changshi Lao,[1] Zhong Lin Wang[1,2,3]*

A previously unknown rigid helical structure of zinc oxide consisting of a superlattice-structured nanobelt was formed spontaneously in a vapor-solid growth process. Starting from a single-crystal stiff nanoribbon dominated by the c-plane polar surfaces, an abrupt structural transformation into the superlattice-structured nanobelt led to the formation of a uniform nanohelix due to a rigid lattice rotation or twisting. The nanohelix was made of two types of alternating and periodically distributed long crystal stripes, which were oriented with their c axes perpendicular to each other. The nanohelix terminated by transforming into a single-crystal nanobelt dominated by nonpolar (01$\bar{1}$0) surfaces. The nanohelix could be manipulated, and its elastic properties were measured, which suggests possible uses in electromechanically coupled sensors, transducers, and resonators.

Helical structures have been observed for a number of inorganic materials. For example, carbon nanotube coils (1) are created when

paired pentagon-heptagon atomic rings arrange themselves periodically within the hexagonal carbon network (2). Formation of nanosprings

of amorphous silica has been thought to occur through a perturbation during the growth of a straight nanowire (3). Helical structures of SiC are proposed to be a screw-dislocation-driven growth process (4). By using the difference in surface stress on the two surfaces, rings and tubes of strained bilayer thin films, such as Si/SiGe, have been made (5).

Zinc oxide, an important semiconductive and piezoelectric material with several applications in optoelectronics and sensing, can form single-crystal nanosprings (6), nanorings (7), and nanobows (8) by bending or folding polar-surface-dominated nanobelts. These growth processes are dominated by a minimization of the electrostatic energy contributed by the cation- and anion-terminated surfaces of this ionic material, and their shapes are determined by balancing the electrostatic energy with the elastic deformation energy. The structural configurations of ZnO nanoobjects (9) could form the basis for fabricating nanoscale sensors, transducers, and resonators.

We now report a distinctive helical structure of ZnO made of a superlattice-structured nanobelt that is formed spontaneously in a solid-vapor growth process. The superlattice nanobelt is a periodic, coherent, epitaxial, and parallel growth of two alternating nanostripes of ZnO crystals oriented with their c axes perpendicular to each other. The role played by polar surfaces in the structural transformation is demonstrated, and the nanohelix has been manipulated by atomic force microscopy (AFM) to measure its elastic properties.

The ZnO nanohelices were grown with high reproducibility by a vapor-solid process (10) and by using temperature to control growth kinetics. The experimental setup consists of a horizontal high-temperature tube furnace, an alumina tube, a rotary pump system, and a gas-controlling system. First, 2 g of commercial ZnO powder (Alfa Aesar, Ward Hill, MA) were compacted, loaded into an alumina boat, and positioned at the center of the alumina tube as the source material. The system was prepumped to $\sim 2 \times 10^{-2}$ mbar, and the ramp rate was controlled at 20°C/min to 25°C/min when the temperature was raised from room temperature to 800°C. The furnace was then held at 800°C for 20 min, and the temperature was ramped at 20°C/min from 800° to 1400°C. When the temperature reached 1000°C, argon was introduced as a carrier gas to raise the pressure from $\sim 2 \times 10^{-2}$ mbar to the desired synthesis pressure of 200 to 250 mbar within ~ 2.5 min. The solid-vapor deposition was carried out at 1400°C for ~ 2

[1]School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, GA 30332–0245, USA. [2]College of Engineering, Peking University, Beijing, China. [3]National Center for Nanoscience and Technology of China, Beijing 100080, China.

*To whom correspondence should be addressed. E-mail: zhong.wang@mse.gatech.edu

hours under a pressure of 200 to 250 mbar. The argon carrier gas was kept at a flow rate of 50 sccm (standard cubic centimeters per minute). The as-grown nanohelices of ZnO were deposited onto a polycrystalline Al_2O_3 substrate at a local temperature of 700° to 800°C and characterized by a variety of microscopes, including scanning electron microscopy (SEM) [LEO 1530 FEG (Oberkochen, Germany) at 5 and 10 kV], high-resolution transmission electron microscopy (TEM) [Hitachi HF-2000 at 200 kV and JEOL 4000EX (Tokyo, Japan) at 400 kV], AFM [Molecular Force Probe MFP-3D from Asylum Research (Santa Barbara, CA)], and focused ion beam microscopy (FIB) (FEI NOVA 200).

The as-synthesized sample was composed of up to 10% of freestanding nanohelices (fig. S1). High-resolution SEM images show the high degree of uniformity of the nanohelices, which can be left- (Fig. 1, A and C) or right-handed (Fig. 1, B and D). Their population ratio is nearly 1:1 (fig. S2). As shown in Fig. 1, the nanohelix diameters, widths, and pitch distances range between 300 to 700 nm, 100 to 500 nm, and 500 to 2500 nm, respectively (11). The length of the nanohelix can be as long as 100 μm. Energy dispersive x-ray spectroscopy analysis indicates that the nanorings are ZnO without detectable impurity.

Low-magnification TEM images of the nanohelices (Fig. 2, A and B) revealed the crystalline structure and winding shape of the nanobelts. The electron diffraction (ED) pattern, acquired from a single period of a nanohelix, shows mirror symmetry. This mirror symmetry is caused by superposition of several ZnO diffraction patterns rotated around the axial direction of the nanohelix axial direction of the nanohelix. A dark-field TEM image recorded from a segment of a nanobelt (Fig. 2C) shows that the nanobelt has a periodic superlattice structure across its entire width, with a uniform periodicity of ~ 3.5 nm. With a 5° offset, the stripes are nearly parallel to the nanobelt growth direction and run along the length of the nanohelix (fig. S3).

From crystallography, the wurtzite-structured ZnO crystal is described schematically as a number of alternating planes of tetrahedrally coordinated O^{2-} and Zn^{2+} ions stacked alternatively along the c axis. The oppositely charged ions produce positively charged (0001)-Zn^{2+} and negatively charged $(000\bar{1})$-O^{2-} polar surfaces (c planes) (7, 9). High-resolution TEM revealed the superlattice structure of the nanobelt with a periodic alternating arrangement of two types of stripes (Fig. 2D), which are oriented epitaxially in the following orientations: Stripe I has a growth direction of $[1\bar{1}00]$, with top and bottom surfaces $\pm(0001)$ (polar surfaces); stripe II has a growth direction of $[2\bar{1}\bar{1}2]$, with top and bottom surfaces $\pm(01\bar{1}0)$ (nonpolar surfaces). The corresponding ED pattern of the nanobelt

shows that the growth directions of the two types of stripes have a small rotation of $\alpha \approx 4°$ (Fig. 2E), indicating that the adjacent stripes rotate in the plane of the nanobelt for the same magnitude of the angle when forming the superlattice. The angles between the ED spots also slightly deviate from the expected result of a large crystal and suggest rigid lattice twist in some atomic planes. The two stripes are coherent by aligning $(0\bar{1}10)$ of stripe I with (0002) of stripe II at a lattice mismatch of 2.9% and $(\bar{2}110)$ of stripe I with $(\bar{2}110)$ of stripe II without lattice mismatch, so the interface energy is likely to be low. Viewing from the projected direction of the incident electron beam, we aligned the two stripes by facing $(\bar{1}\bar{1}20)$ of stripe I parallel to the $(\bar{2}116)$ of stripe II and aligned the $[0001]$ of stripe I with $[01\bar{1}0]$ of stripe II. The true interface is inclined with respect to the direction of the electron beam, so that the contrast at the
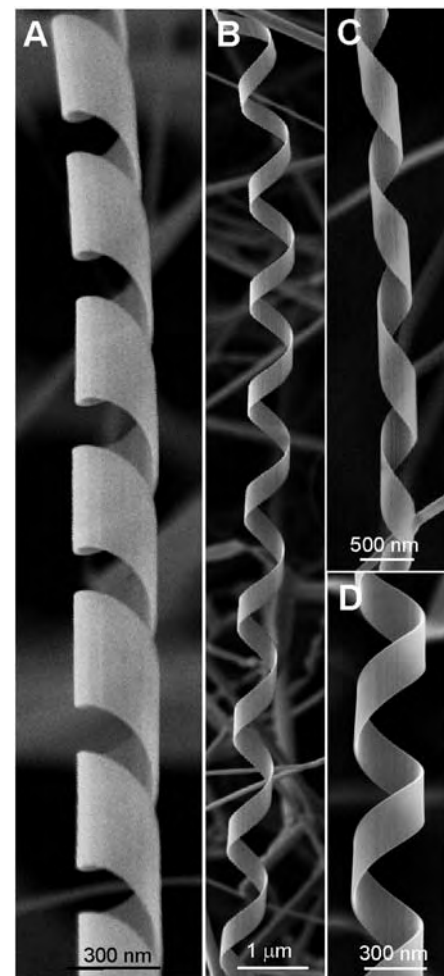

Fig. 1. Morphology of nanohelices. (**A** to **C**) Typical SEM images of the left- and right-handed ZnO nanohelices. (**D**) High-magnification SEM image of a right-handed ZnO nanohelix, showing a uniform and perfect shape. All of the nanohelices are directly received from the as-synthesized sample, and they are freestanding on the substrate surface.

interface in the high-resolution TEM image is not very sharp (Fig. 2F). On the basis of the orientation relation presented above, the experimental image can be simulated by using dynamic ED theory. The result is inserted in Fig. 2F (*11*), showing quantitative agreement with the observation. Thus, the interface for stripe I is likely to be $(\bar{1}\bar{1}22)$ (*12*).

The initiation of the nanohelix is a direct result of superlattice formation. From an SEM image of a nanohelix with both ends (Fig. 3A), a straight stiff nanoribbon is directly connected to the starting point of a nanohelix. TEM examination shows that the joining point between the nanoribbon and the nanohelix is the place where the superlattice is created (Fig. 3B). The stiff nanoribbon is a single crystal object dominated by large *c*-plane surfaces (Fig. 3C). The superlattice is formed spontaneously and abruptly across the entire width of the nanoribbon. A striking feature is that the width of the nanobelt increases and its thickness decreases significantly after transforming into the superlattice structure (Fig. 3B and fig. S4). The stripes hold the same orientation relation as defined in Fig. 2E [see the selected-area ED (SAED) pattern inserted in Fig. 3C]. An enlarged local area shows the coexistence and spontaneous formation of two periodically arranged, alternating stripes at the transverse interface where the structural transformation occurred abruptly (Fig. 3D). At the transverse interface, the lattice mismatch of 2.9% between $(0\bar{1}10)$ of the nanoribbon with (0002) of stripe II should create interface mismatch edge dislocations with a Burgers vector of $[0\bar{1}10]$, which are expected to be separated by 3.45 nm. In fact, the periodicity of the superlattice is ~3.5 nm, and mismatch edge dislocations have been identified at the turning point between the nanoribbon and the superlattice (circled areas in Fig. 3D). The site of the mismatch dislocation could be the nucleation site of stripe II. The superior uniformity in periodicity of the superlattice is most likely to be a result of the lattice mismatch as defined by crystal structure, and rigid crystal twisting gives rise to the strain contrast observed in Fig. 3B and even in the high-resolution TEM image in Fig. 3D.

The termination of the nanohelix growth is caused by a complete conversion of the remaining polar surfaces into nonpolar surfaces and an increase in nanoribbon thickness. From the TEM image and the corresponding SAED patterns recorded from a single nanohelix (Fig. 3E), the formation of the nanohelix is a result of transforming a fully *c*-plane-dominated stiff nanoribbon (starting point) into a superlattice-structured, partially polar-surface-dominated flexible nanobelt (middle section). The starting stiff nanobelt is thick (~ 40 nm) and straight (fig. S4). After transforming into the superlattice-structured nanobelt, the thickness of the nanobelt is greatly reduced to less than

Fig. 2. Crystal structure of the nanohelix. **(A)** Typical low-magnification TEM image of a ZnO nanohelix, showing its structural uniformity. A straight nanobelt was enclosed inside the helix during the growth. **(B)** Low-magnification TEM image of a ZnO nanohelix with a larger pitch-to-diameter ratio. The selected-area ED pattern (SAEDP, inset) is from a full turn of the helix. **(C)** Dark-field TEM image from a segment of a nanohelix, showing that the nanobelt that coils into a helix is composed of uniformly parallel, longitudinal, and alternatively distributed stripes at a periodicity of ~3.5 nm across its entire width. The edge at the right-hand side is the edge of the nanobelt. **(D** and **E)** High-magnification TEM image and the corresponding SAEDP of a ZnO nanohelix with the incident beam perpendicular to the surface of the nanobelt, respectively, showing the lattice structure of the two alternating stripes. The SAEDP is composed of two sets of fundamental patterns, labeled and indexed in red for stripe I and yellow for stripe II. A careful examination of the image indicates that the true interface between the stripes is not edge-on in reference to the incident electron beam but at a relatively large angle. **(F)** Enlarged high-resolution TEM image showing the interface between the two adjacent stripes. The area within the dotted line is a simulated image using the dynamic electron diffraction theory. The interface proposed here is $(\bar{1}\bar{1}22)$ for stripe I, which is inclined with respect to the incident electron beam at an angle of 32°.
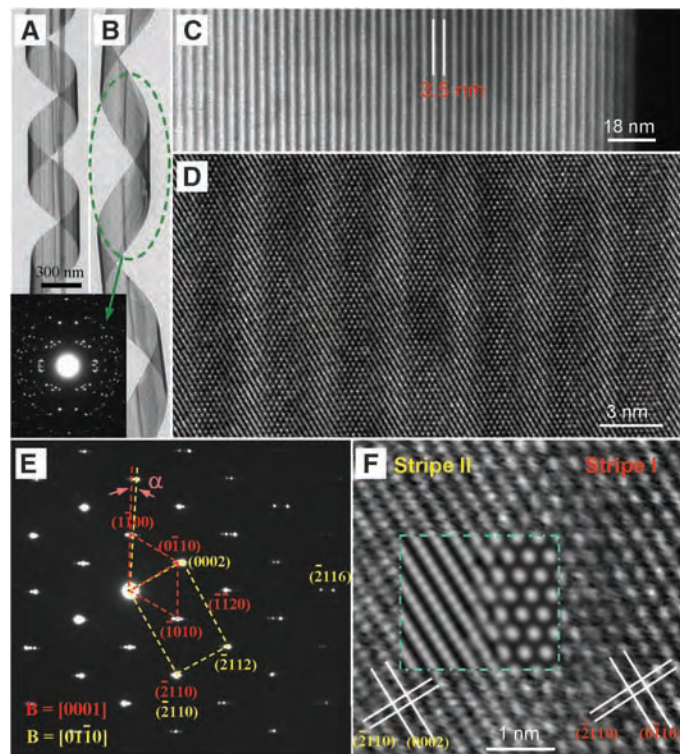
20 nm, and its width is substantially increased (fig. S4). Lastly, the ending of the nanohelix is a result of abruptly transforming the partially *c*-plane-dominated, superlattice-structured nanobelt into a single-crystal nanoribbon dominated by nonpolar surfaces at the finishing end (Fig. 3E), where the corresponding dark-field TEM image and the high-resolution TEM image (Fig. 3F) indicate that the top and bottom surfaces of the nanoribbon are the nonpolar $\pm(01\bar{1}0)$ surfaces. The thickness of the nanoribbon increases significantly as well after the transformation.

The formation of the nanohelix is a rigid structural alteration caused by the alternating stripes in the superlattice, and the formation of a superlattice is likely to be a polar charge–induced structural transformation (see the model in Fig. 3G). In the synthesis, a pre-pumping to 2×10^{-2} mbar creates the condition for growing polar surface–dominated nanoribbons (*13*). The polar charges are preserved on the surface during growth under conditions of high temperature and inert environment (*14*). As reported previously (*6*, *7*), the existence of the polar surfaces results in the formation of ZnO nanosprings and nanorings. If the width of the nanoribbon is rather large

(Fig. 3B), the electrostatic energy of the polar surfaces diverges with the extension of its length. A possible way of reducing the electrostatic energy is to transform the polar surface into a nonpolar surface, provided the interface mismatch energy is reasonably low, forming superlattice-structured stripes that are nearly parallel to the growth direction of the nanobelt. The introduction of argon gas at a later stage during the growth may be the initiation of superlattice formation. The switching in the *c* axis may indicate the formation of piezo-electric domains. The width of the nanobelt increases after the structural transformation from a single-crystal, *c*-plane-dominated stiff nanoribbon (in green) into a superlattice-structured, flexible nanobelt; the stripes are not exactly parallel to the growth direction of the nanobelt, and a small in-plane rotation for an angle of α exists between the growth directions of the two adjacent stripes, which induces the geometrical curvature required for rigidly twisting its shape. The polar charges remaining on the surfaces of stripe I also help to bend the nanobelt into a ring structure, especially with the decrease of the nanobelt thickness (*15*). Thus, the accumulation of rigid structural rotation or twist across the width of
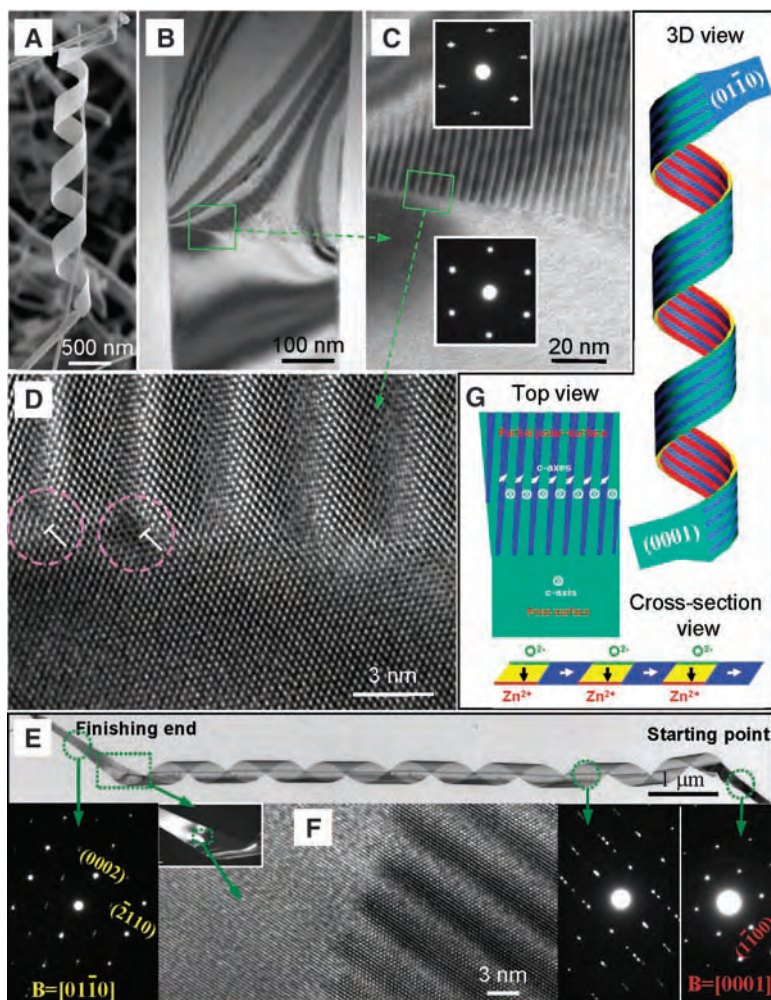
Fig. 3. Initiation and formation process of a nanohelix. (**A**) Low-magnification SEM image of a ZnO nanohelix showing its starting point and finishing end. (**B**) Low-magnification TEM image from the starting point of a nanohelix, showing a distinct and sharp contrast change across a transverse line and an increase in nanobelt width. (**C**) Enlargement of the boxed area in (B) and the corresponding diffraction patterns from the single crystal region (bottom inset) and the superlattice-structured region (top inset). The bottom segment is a *c* plane–dominated single-crystal stiff nanoribbon, whereas the top segment is a superlattice-structured flexible nanobelt, and the corresponding electron diffraction pattern is the same as the one shown in Fig. 2E. (**D**) High-magnification TEM image recorded from the interfacial region of the segments, showing a clear cutting line from where the structural transformation occurs across the entire width of the nanobelt. Mismatch edge dislocations are identified in the circled regions and directly linked to stripe II. (**E**) Low-magnification TEM image of a nanohelix showing its entire structure from beginning to end. Three representative SAEDPs are recorded from areas selected from the starting end, the middle section, and the finishing end. The starting point is a nanoribbon oriented along [0001], which is fully dominated by the *c*-plane polar surfaces; the middle section of the nanohelix is superlattice-structured, whereas the finishing end is a single-crystal nanoribbon that is dominated by nonpolar surfaces according to the dark-field TEM image (inset), [01$\bar{1}$0] diffraction pattern, and (**F**) high-resolution TEM image recorded from the area, where a distinct and clear-cut transformation from the superlattice-structured nanobelt to an *a*-plane-dominated nanoribbon is apparent. (**G**) Schematic models showing the top view and cross-sectional view of the nanobelt for illustrating the corresponding directions of the *c* axes and the distribution of polar charges (in green for O^{2-} and red for Zn^{2+}) and nonpolar surfaces (in blue). A three-dimensional (3D) structure model of a nanohelix, showing its initiating point and finishing end. The periodicity of the superlattices may result in the formation of periodic piezoelectric domains.

the superlattice nanobelt and a continuous growth of the stripes along the length of the nanobelt result in the formation of a helical structure, which could be left-handed or right-handed. The winding of the nanobelt is terminated after the nanobelt fully transforms from a partial-polar-surface-dominated and superlattice-structured nanoobject into a single-crystal nanoribbon dominated by nonpolar surface (in blue), which is likely to be caused by a change in growth kinetics. The single-crystal nanoribbon is a straight structure without rigid lattice alteration or surface polar charges; thus, it will not continue to wind but will extend as a straight ribbon.

Manipulation of a nanohelix with AFM and FIB shows that it has a rigid structure. A nanohelix has been sectioned by FIB into a short segment of one period (Fig. 4A), but the shape of the segment is still preserved and shows no visible relaxation. A nanohelix has also been cut by an AFM tip at two consecutive turns (Fig. 4C), and the preservation of the rigid shape by a single-turn helix after cutting suggests little, if any, stored elastic energy (*16*).

The transverse spring constant of the nanohelix, a quantity for describing the stiffness of

Fig. 4. Manipulation of a nanohelix and measurement of its elastic properties. (**A**) SEM image of a nanohelix that was cut by a FIB microscope shows the preservation of the rigid geometrical shape after the cut, which suggests there is minimal stored elastic energy. (**B**) AFM contact mode measurement of the sensitivity (*S*) of the photodetector for the silicon substrate and the four turns of a ZnO nanohelix, as indicated by 1 to 4 in the inserted AFM image. The sensitivity *S* of the photodetector is the inverse of the slope of the force-displacement (*F-Z*) curve, which is directly related to the elastic property. The

measurement was done by positioning the tip on the surface of the upmost middle point of the turn of the helix. Each dot represents one measurement. The sensitivity S_{Si} for silicon is substantially lower than that for the ZnO nanohelix (S_{Helix}). (**C**) Fracture of the nanohelix at two consecutive turns of the nanohelix by the AFM tip, showing a preservation of helical shape in the sectioned region.
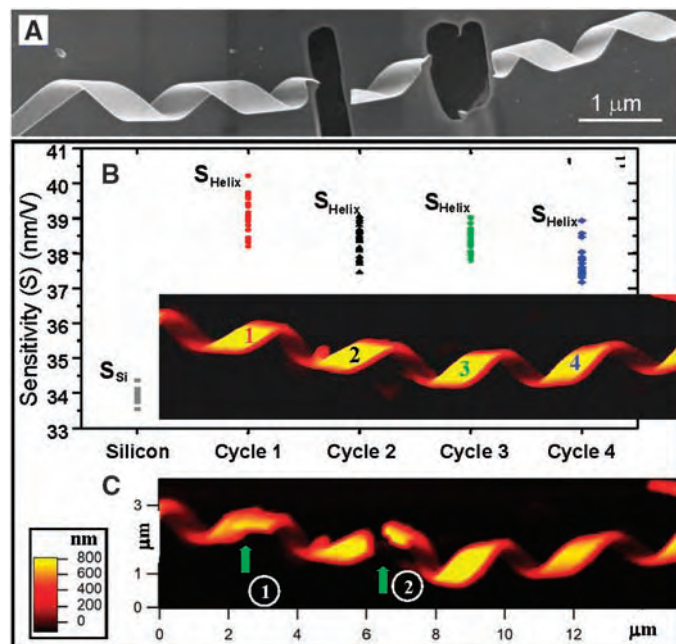
the nanohelix perpendicular to its axial direction, has been measured by AFM with use of the force-displacement (F-Z) curve (*17*). Through the static compression of a nanohelix lying on a silicon substrate with an AFM tip (Fig. 4B) (fig. S5) and careful calibration of the sensitivity, S, of the photodetector (*11, 18*), which is the inverse of the slope of the F-Z curve (fig. S6), the transverse spring constant of the nanohelix, K_{Helix}, was found to be (*11*) 3.9, 4.6, 4.5, and 5.3 N/m for one to four turns, respectively, of the nanohelix as labeled in Fig. 4B. The measured K_{Helix} is near the theoretically estimated transverse spring constant of 4.2 N/m (*11*), and so the elastic modulus, E, values of the nanohelix derived from the measured spring constant (*11*) are 42, 49, 48, and 57 GPa for the four turns, respectively, which agrees well with the elastic modulus measured for straight ZnO nanobelts by mechanical resonance with the use of in situ TEM (*19*).

The perfect helix we observed is of great interest not only for understanding the growth behavior of polar-surface–driven growth processes in the wurtzite system, such as GaN, AlN, and InN, but also for investigating fundamental physics and optical phenomena. The piezoelectric and semiconducting properties of ZnO suggest that the nanohelix could be a fundamental unit for investigating electromechanically coupled nanodevices by using the superlattice piezoelectric domains. The nanohelix is likely to have important applications in sensors, transducers, resonators, and photonics.

References and Notes

1. S. Amelinckx *et al.*, *Science* **265**, 635 (1994).
2. R. P. Gao, Z. L. Wang, S. S. Fan, *J. Phys. Chem. B* **104**, 1227 (2000).
3. H. F. Zhang, C. M. Wang, E. C. Buck, L. S. Wang, *Nano Lett.* **3**, 577 (2003).
4. H. F. Zhang, C. M. Wang, L. S. Wang, *Nano Lett.* **2**, 941 (2002).
5. O. G. Schmidt, K. Eberl, *Nature* **410**, 168 (2001).
6. X. Y. Kong, Z. L. Wang, *Nano Lett.* **3**, 1625 (2003).
7. X. Y. Kong, Y. Ding, R. Yang, Z. L. Wang, *Science* **303**, 1348 (2004).
8. W. L. Hughes, Z. L. Wang, *J. Am. Chem. Soc.* **126**, 6703 (2004).
9. Z. L. Wang *et al.*, *Adv. Funct. Mater.* **14**, 944 (2004).
10. Z. W. Pan, Z. R. Dai, Z. L. Wang, *Science* **291**, 1947 (2001).
11. See Materials and Methods provided on *Science* Online.
12. The {1̄1̄22} are typical twin planes for ZnO (*20*).
13. P. X. Gao, Z. L. Wang, *Small*, in press; published online (http://dx.doi.org/10.1002/smll.200500165).
14. The growth of the nanostructure occurred at a local temperature of ~700° to 800°C in an argon atmosphere, under which the possibility of neutralizing the surface polar charges by adsorbing foreign molecules is unlikely, provided the growth occurred fairly quickly. This result is consistent with our previous studies [see (*7–9*)].
15. From the energy point of view, a polar-surface–dominated nanobelt tends to fold itself into a ring or spring for reducing electrostatic energy (*6*). The nanoring is stable if the ratio between nanobelt thickness, t, and the radius, R, of the nanoring is smaller than ~3% (*8*). Nanorings with $t/R > 3\%$ are energetically unfavorable to form.
16. The aforementioned data rule out the possibility that the nanohelix is formed by bilayer surface strain (*5*) for two reasons. First, there is no bilayer strained structure in our system; and second, the difference in surface stress on the (0001)-Zn and (0001̄)-O of ZnO, if any, has negligible effect on the formation of nanorings or nanobows (*8*). The nanohelix is likely to be formed by two processes: the rigid structural alteration due to the formation of the superlattice and the spontaneous polar-charge–induced bending (*6*). However, t of the nanobelt is ~20 nm, and typical R of the nanohelix is ~150 to 350 nm (Fig. 1). Thus the t/R is ~6 to 13%, which is much larger than the threshold value (~3%) permitted for forming a nanoring by electrostatic polar charges (*15*). Also, considering the great reduction in area of the polar surfaces (for ~50%) after forming the superlattice as well as the helical shape of the structure, the role played by polar charges in forming a helix is dramatically reduced. Therefore, the dominant mechanism for forming the nanohelix is by rigid lattice rotation and twisting as a result of superlattice formation.
17. E. W. Wong, P. E. Sheehan, C. M. Lieber, *Science* **277**, 1971 (1997).
18. A. Volodin *et al.*, *Phys. Rev. Lett.* **84**, 3342 (2000).
19. X. D. Bai, P. X. Gao, Z. L. Wang, E. G. Wang, *Appl. Phys. Lett.* **82**, 4806 (2003).
20. Y. Dai, Y. Zhang, Z. L. Wang, *Solid State Commun.* **126**, 629 (2003).
21. Thanks to support from NSF, the NASA Vehicle Systems Program and Department of Defense Research and Engineering, the Defense Advanced Research Projects Agency, and the Chinese Academy of Sciences.

Simulated Quantum Computation of Molecular Energies

Alán Aspuru-Guzik,[1]*† Anthony D. Dutoi,[1]* Peter J. Love,[2] Martin Head-Gordon[1,3]

The calculation time for the energy of atoms and molecules scales exponentially with system size on a classical computer but polynomially using quantum algorithms. We demonstrate that such algorithms can be applied to problems of chemical interest using modest numbers of quantum bits. Calculations of the water and lithium hydride molecular ground-state energies have been carried out on a quantum computer simulator using a recursive phase-estimation algorithm. The recursive algorithm reduces the number of quantum bits required for the readout register from about 20 to 4. Mappings of the molecular wave function to the quantum bits are described. An adiabatic method for the preparation of a good approximate ground-state wave function is described and demonstrated for a stretched hydrogen molecule. The number of quantum bits required scales linearly with the number of basis functions, and the number of gates required grows polynomially with the number of quantum bits.

Feynman observed that simulation of quantum systems might be easier on computers using quantum bits (qubits) (*1*). The subsequent development of quantum algorithms has made this observation concrete (*2–6*). On classical computers, resource requirements for complete simulation of the time-independent Schrödinger equation scale exponentially with the number of atoms in a molecule, limiting such full configuration interaction (FCI) calculations to diatomic and triatomic molecules (*7*). Computational quantum chemistry is therefore based on approximate methods that often succeed in predicting chemical properties for larger systems, but their level of accuracy varies with the nature of the species, making more complete methods desirable (*8*).

Could quantum computation offer a new way forward for exact methods? Despite the formal promise, it has not been demonstrated that quantum algorithms can compute quantities of chemical importance for real molecular systems to the requisite accuracy. We address this issue by classically simulating quantum computations of the FCI ground-state energies of two small molecules. Although the basis sets used are small, the energies are obtained to the precision necessary for chemistry. Absolute molecular energies must be computed to a precision (greater than six decimal places) that reflects the smaller energy differences observed in chemical reactions (~0.1 kcal/mol). These simulations show that quantum computers of tens to hundreds of qubits can match and exceed the capabilities of classical FCI calculations.

A molecular ground-state energy is the lowest eigenvalue of a time-independent Schrödinger equation. The phase-estimation algorithm (PEA) of Abrams and Lloyd (*3, 4*) can

---
[1]Department of Chemistry, University of California, Berkeley, CA, USA. [2]D-Wave Systems, Inc., 4401 Still Creek Drive, Suite 100, Burnaby, BC V5C 6G9, Canada. [3]Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: alan@aspuru.com

be used to obtain eigenvalues of Hermitian operators; we address issues concerning its implementation for molecular Hamiltonians. The molecular ground-state wave function $|\Psi\rangle$ is represented on a qubit register **S** (state). Another register **R** (readout) is used to store intermediate information and to obtain the Hamiltonian eigenvalue E. The Hamiltonian $\hat{H}$ is used to generate a unitary operator $\hat{U}$, with E mapped to the phase of its eigenvalue $e^{i2\pi\phi}$.

$$\hat{U}|\Psi\rangle = e^{i\hat{H}\tau}|\Psi\rangle = e^{i2\pi\phi}|\Psi\rangle; \quad E = 2\pi\phi/\tau \quad (1)$$

Through repeated controlled action of powers of $\hat{U}$, the computer is placed in the state

$$|\mathbf{R}\rangle \otimes |\mathbf{S}\rangle = \left(\sum_n e^{(i2\pi\phi)n}|n\rangle\right) \otimes |\Psi\rangle \quad (2)$$

The summation index n enumerates the basis states of **R** according to their bit-string value. The quantum inverse Fourier transform is then applied to **R** to obtain an approximation to ϕ written in binary to **R**. The procedure is related to the Fourier transform of the time dependence of an eigenstate to obtain its eigenenergy. By using polynomially scaling classical approximation methods, an initial estimate of E can be obtained to choose τ such that $0 \leq (\phi \approx 1/2) < 1$.

We address four separate issues. First, we show how standard chemical basis sets can be used for representations of the wave function on **S**. Second, although the size of **R** relative to **S** will be marginal in the large-system limit, this initial overhead (20 qubits for a chemically meaningful result) presently represents a substantial impediment to both classical simulation and actual implementation of the algorithm. We show how a modification of the PEA makes it possible to perform a sequence of computations with a smaller register, such that the precision of the result obtained is independent of the size of **R**. Third, the algorithm requires that any estimated ground state has a large overlap with the actual eigenstate. We show how a good estimate of the ground-state wave function may be prepared adiabatically from a crude starting point. Finally, $\hat{U}$ must be represented in a number of quantum gates that scales polynomially with the size of the system, and we give such bounds.

Any implementation of a quantum-simulation algorithm requires a mapping from the system wave function to the state of the qubits. Basis-set methods of quantum chemistry often represent many-particle molecular wave functions in terms of single-particle atomic orbitals. The number of orbitals in a basis set is proportional to the number of atoms in a molecule. The molecular wave function may be represented by a state of **S** in two basic ways. In the direct mapping, each qubit represents the fermionic occupation state of a particular atomic orbital,

occupied or not. In this approach, a Fock space of the molecular system is mapped onto the Hilbert space of the qubits. This mapping is the least efficient but has advantages discussed later. In the more efficient compact mapping, only a subspace of the Fock space with fixed electron number is mapped onto the qubits. The states of the simulated system and of the qubit system are simply enumerated and equated. Furthermore, one could choose only a subspace of the fixed-particle-number space. The compact mapping with restriction to a spin-state subspace is the most economical mapping considered in this work. Figure 1 shows that the number of qubits required for both the compact and direct mappings scales linearly with the number of basis functions. Also shown are the qubit requirements for specific molecules with different basis sets and mappings. More extensive qubit estimates for computations on H_2O are given in Table 1, including restriction to the singlet-spin subspace.

In this work, a modified PEA was carried out, which uses a relatively small number of qubits in **R** (as few as four for stability). This implementation allows more of the qubits to be devoted to information about the system and decreases the number of consecutive coherent quantum gates necessary. This procedure can be interpreted as making continually better estimates of a reference energy. The Hamiltonian is then shifted by the current reference energy and an estimate of the deviation of the actual energy from the reference is computed. The reference energy is then updated, and the procedure is repeated until the desired precision is obtained.

The algorithm at iteration k is illustrated in Fig. 2A. In iteration zero, we set $\hat{V}_0 = \hat{U}$ and
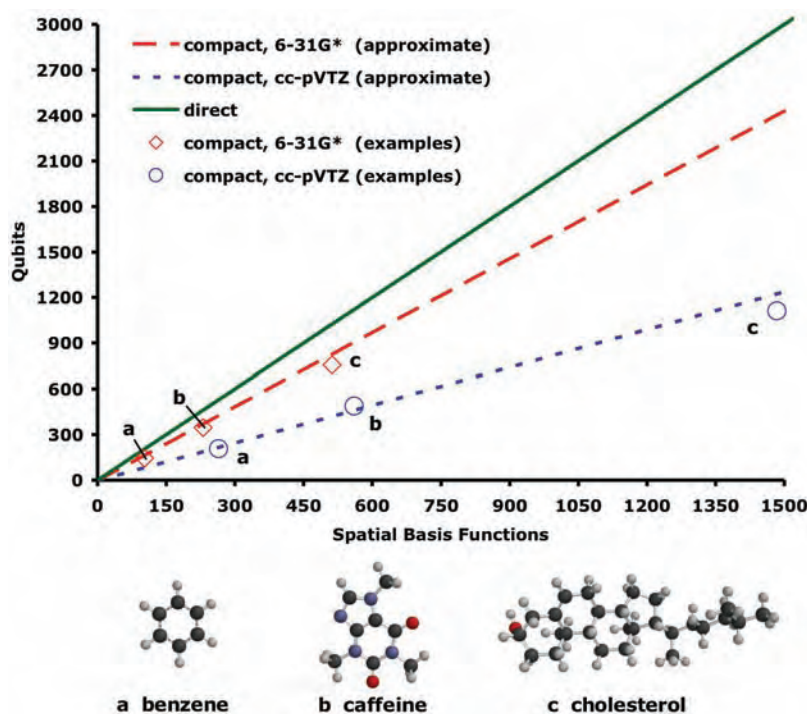
Fig. 1. Qubit requirements versus basis size. The number of qubits required to store the wave function of a molecule is shown as a function of the number of basis functions for different mappings. For the compact mapping, the qubit requirement depends also on the ratio of number of electrons to basis functions, which is relatively constant for a given basis set; although the higher quality cc-pVTZ basis is more economical per basis function, a molecule in this basis uses substantially more functions than with the 6-31G* basis. The qubits required for specific molecules and basis sets are also shown.

Table 1. Qubit requirements for computations on water. The number of qubits needed to store the wave function of water is given for various basis sets and system-qubit mappings, including restriction to the singlet-spin subspace.

Water	Basis set (number of functions)		
Mapping	STO-3G (7)	6-31G* (19)	cc-pVTZ (58)
Compact (singlets)	8	25	42
Compact	10	29	47
Direct	14	38	116

Fig. 2. Recursive PEA circuit and output. (A) The quantum circuit for the recursive phase-estimation algorithm is illustrated. k iterations are required to obtain k bits of a phase ϕ that represents the molecular energy. QFT^+ represents the quantum inverse Fourier transform and Hd is a Hadamard gate; the dial symbols represent measurement. (B) Output probabilities for obtaining the first eight bits of ϕ in the water calculation are shown. The abscissa is scaled to be in terms of molecular energy, and the ordinate is probability.

$$\hat{V}_k = \left[\exp(-i2\pi\phi_{k-1}) \hat{V}_{k-1} \right]^2$$

Fig. 3. ASP evolution of ground-state overlap and excitation gap. (A) Time evolution of the squared overlap of the wave function $|\Psi_{ASP}\rangle$ with the exact ground state $|\Psi\rangle$ during adiabatic state preparation is shown. The system is the hydrogen molecule at different nuclear separations r; time was divided into 1000 steps in all 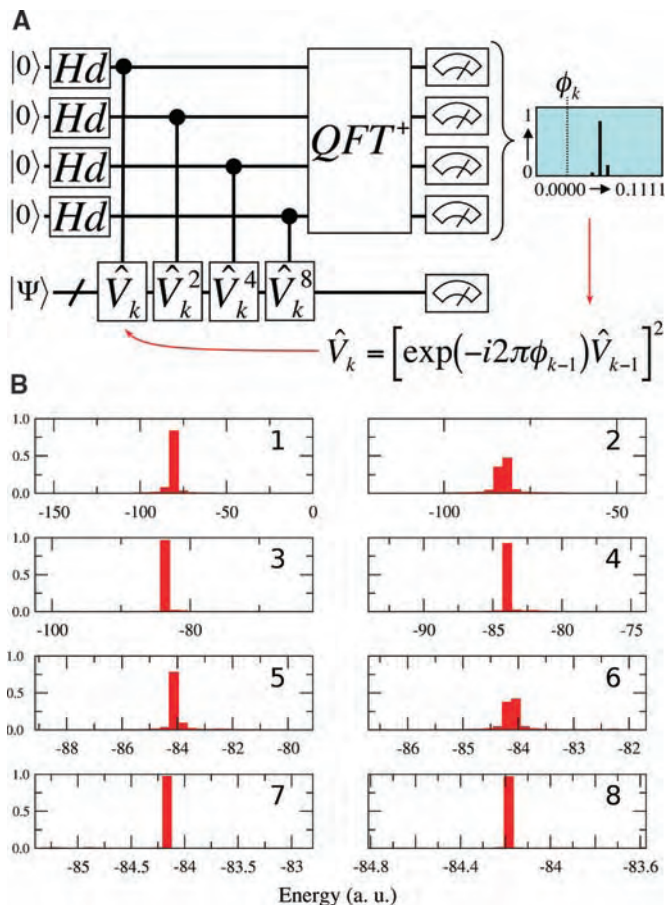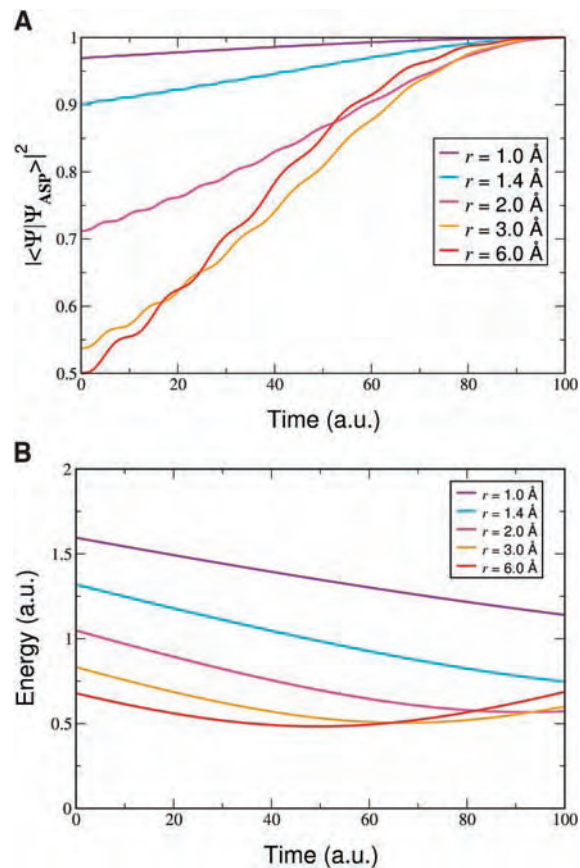cases. (B) Time evolution of the singlet ground- to first-excited-state energy gap of the Hamiltonian used along the adiabatic path is shown.

perform a four-qubit PEA on $\hat{V}_0$. This estimates ϕ on the interval zero to unity with a precision of 1/16. We use this estimate to construct a shift ϕ_0, which is a lower bound on ϕ. We apply this shift and repeat the four-qubit PEA using the new operator $\hat{V}_1 = \left[e^{-i2\pi\phi_0} \hat{V}_0 \right]^2$. This determines the remainder of ϕ above the previous lower bound on an interval representing half of the previous interval. In each subsequent iteration k, we construct a similarly modified operator $\hat{V}_k$ and shift ϕ_k. By choosing a ϕ_k that is one-fourth lower than the phase of the $\hat{V}_k$ eigenvalue estimate, we ensure that the phase of the $\hat{V}_{k+1}$ eigenvalue is approximately centered on the interval zero to unity. In each iteration, we therefore obtain one additional bit of ϕ, as shown in Fig. 2B for a calculation on H_2O.

To demonstrate the usefulness of the recursive procedure, we carried out calculations on H_2O and LiH. For H_2O, we used the minimal STO-3G basis set, yielding 196 singlet-spin configurations; there are 1210 such configurations for LiH in the larger 6-31G basis. This required 8 and 11 qubits, respectively, for the compact mapping of the singlet subspace. Register **S** was initialized to the Hartree-Fock (HF) wave function in both cases. After 20 iterations, the electronic energy obtained for H_2O [–84.203663 atomic units (a.u.)] matched the Hamiltonian diagonalization energy (–84.203665 a.u.). The LiH calculation (–9.1228936 a.u.) matched diagonalization (–9.1228934 a.u.) to the same number of significant digits. The precision is good enough for almost all chemical purposes. The discrepancy between the PEA and diagonalization is attributed to error in matrix exponentiation to form $\hat{U}$ from $\hat{H}$.

In the simulations described above, the approximation to the ground-state wave function was the HF state $|\Psi^{HF}\rangle$. The probability of observing the exact ground state $|\Psi\rangle$, and hence the success of the PEA, is then proportional to $|\langle\Psi|\Psi^{HF}\rangle|^2$. However, it is known for some cases, such as molecules close to the dissociation limit or in the limit of large system size, that the HF wave function has vanishing overlap with the ground state (9). The overlap of the initially prepared state with the exact state can be systematically improved by an adiabatic-state-preparation (ASP) algorithm, relying on the adiabatic theorem (10–12). The theorem states that a system will remain in its ground state if the Hamiltonian is changed slowly enough. Our Hamiltonian is changed slowly by discretized linear interpolation from the trivial HF case to the FCI operator. The efficiency is governed by how rapidly the Hamiltonian may be varied, which is determined by the gap between ground-state and first-excited-state energies along the path (11). In the case of quantum chemistry problems, lower bounds on this gap may be estimated with conventional methods.

The path $\hat{H}^{HF} \to \hat{H}$ is chosen by defining $\hat{H}^{HF}$ to have all matrix elements equal to zero, except the first element, namely $H_{1,1}$, which is equal to the HF energy. This yields an initial gap the size of the ground-state mean-field energy, which is very large relative to typical electronic excitations. The ASP method was applied to the H_2 molecule at large separations in the STO-3G basis, for which the squared overlap of the HF wave function with the exact ground state is one half. As evidenced by Fig. 3A, the ASP algorithm prepares states with a high squared overlap for several internuclear distances of the H_2 molecule. Figure 3B plots the relevant gap along the adiabatic path, which is shown for this system to be well-behaved and nonvanishing.

The accuracy and quantum-gate complexity of the algorithm depend on the specific gate decomposition of the unitary operators $\hat{V}_k$, defined above. The factorization of unitary matrices into products of one- and two-qubit elementary gates is the fundamental problem of quantum circuit design. We now demonstrate that the lengths of the gate sequences involved are bounded from above by a polynomial function of the number of qubits.

We analyze the gate complexity of our $\hat{U}$ for the direct mapping of the state. The molecular Hamiltonian is written in second quantized form as

$$\hat{H} = \sum_X \hat{h}_X = \sum_{p,q} \langle p|\hat{T} + \hat{V}_N|q\rangle \, \hat{a}_p^+ \hat{a}_q - \frac{1}{2}\sum_{p,q,r,s} \langle p|\langle q|\hat{V}_e|r\rangle|s\rangle \, \hat{a}_p^+ \hat{a}_q^+ \hat{a}_r \hat{a}_s \qquad (3)$$

where $|p\rangle$ is a one-particle state, $\hat{a}_p$ is its fermionic annihilation operator, and $\hat{T}$, $\hat{V}_N$, and $\hat{V}_e$ are the one-particle kinetic and nuclear-attraction operators and the two-particle electron-repulsion operator, respectively. It has been shown in (2), that for the following approximation to $\hat{U}$

$$e^{i\hat{H}\tau} \approx \left[\prod_X e^{i\hat{h}_X \frac{\tau}{M}}\right]^M \qquad (4)$$

M can always be chosen such that the error is bounded by some preset threshold. The number of gates to implement $\hat{U}$ then scales polynomially with the system size for a given M, under the conditions that the number of terms $\hat{h}_x$ scales polynomially with system size and that each $\hat{h}_x$ acts on a polynomially scaling number of qubits. In our case, these conditions are manifestly fulfilled. The number of terms in the Hamiltonian grows approximately with the fourth power of the number of atoms, and each term involves maximally four basis functions, implying action on at most five qubits in the direct mapping (four qubits in **S** plus a control qubit in **R**). A linear-scaling number of two-qubit operations (similar to qubit swaps) can

account for fermionic antisymmetry in the action of the unitary operator constructed from each $\hat{h}_X$ (13). M is a multiplicative factor in the number of gates. Because the fraction of all pairs of $\hat{h}_X$ terms that do not commute decreases with system size, it is reasonable to assume that M increases polynomially at worst. The advantage of the direct mapping is that, at most, controlled four-qubit unitary operations are required. The number of one- and two-qubit elementary gates required to represent an arbitrary four-qubit gate has been shown to be always less than 400 (14); the structure of a controlled four-qubit unitary operation will allow a decomposition into a similar order of magnitude in the number of gates.

We have found that chemical precision can be achieved with modest qubit requirements for the representation of the wave function and for the readout register. The ASP algorithm has been shown to systematically improve the probability of success of the PEA. Although exponentially difficult on a classical computer, extension to larger molecules requires only linear growth in the number of qubits. The direct mapping for the molecular wave function to the qubit state allows the unitary operator to be decomposed into a number of gates that scales polynomially with system size.

The difficulty of performing quantum-computing simulations is about an order of magnitude greater than conventional FCI. Although possible as experiments, such simulations are not a competitive alternative. To repeat the calculations performed here with a high-quality basis set (cc-pVTZ) would require **S** to consist of 47 or 22 qubits for H_2O or LiH, respectively, using the compact mapping of the full Hilbert space. For most molecules and basis set combinations shown in Fig. 1, an FCI calculation is certainly classically intractable. An FCI calculation for H_2O with cc-pVTZ would be at the edge of what is presently possible. This demonstrates an often-stated conjecture, that quantum simulation algorithms with 30 to 100 qubits will be among the smallest applications of quantum computing that can exceed the limitations of classical computing.

References and Notes

1. R. P. Feynman, *Int. J. Theor. Phys.* **21**, 467 (1982).
2. S. Lloyd, *Science* **273**, 1073 (1996).
3. D. S. Abrams, S. Lloyd, *Phys. Rev. Lett.* **79**, 2586 (1997).
4. D. S. Abrams, S. Lloyd, *Phys. Rev. Lett.* **83**, 5162 (1999).
5. A. Y. Kitaev, *arXiv e-print*; www.arxiv.org/abs/quant-ph/9511026 (1995).
6. R. Cleve, A. Ekert, C. Macciavello, M. Mosca, *Proc. R. Soc. London Ser. A* **454**, 313 (1998).
7. L. Thøgersen, J. Olsen, *Chem. Phys. Lett.* **393**, 36 (2004).
8. T. Helgaker, P. Jorgensen, J. Olsen, *Molecular Electronic-Structure Theory* (Wiley, Sussex, 2002).
9. W. Kohn, *Rev. Mod. Phys.* **71**, 1253 (1999).
10. M. Born, V. Fock, *Zeit. Phys.* **51**, 165 (1928).
11. E. Farhi, J. Goldstone, S. Gutmann, M. Sipser, *arXiv e-print*; www.arxiv.org/abs/quant-ph/0007071 (2000).
12. E. Farhi *et al.*, *Science* **292**, 472 (2001).
13. G. Ortiz, J. E. Gubernatis, E. Knill, R. Laflamme, *Phys. Rev. A.* **64**, 022319 (2001).
14. V. Bergholm, J. Vartainen, M. Möttönnen, M. M. Salomaa, *Phys. Rev. A.* **71**, 052330 (2005).
15. This work was supported by a research grant from D-Wave systems.

Fe-Mg Interdiffusion in (Mg,Fe)SiO₃ Perovskite and Lower Mantle Reequilibration

Christian Holzapfel,[1]*† **David C. Rubie,**[1] **Daniel J. Frost,**[1] **Falko Langenhorst**[2]

Fe-Mg interdiffusion coefficients for (Mg,Fe)SiO₃ perovskite have been measured at pressures of 22 to 26 gigapascals and temperatures between 1973 and 2273 kelvin. Perovskite Fe-Mg interdiffusion is as slow as Si self-diffusion and is orders of magnitude slower than Fe-Mg diffusion in other mantle minerals. Length scales over which chemical heterogeneities can homogenize, throughout the depth range of the lower mantle, are limited to a few meters even on time scales equivalent to the age of Earth. Heterogeneities can therefore only equilibrate chemically when they are stretched and thinned by intense deformation.

The kinetics of many physical and chemical processes in Earth's mantle are controlled by solid-state diffusion (1–5). Thus, understanding and quantifying these processes in Earth requires knowledge of diffusion coefficients for mantle minerals over the range of pressure-temperature conditions encompassed by Earth's mantle. The mineralogy of Earth's lower mantle is dominated by silicate perovskite (~80 volume %).
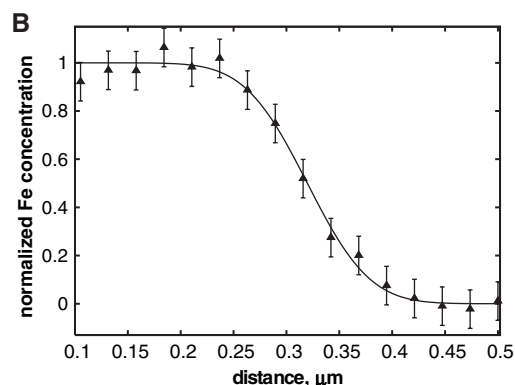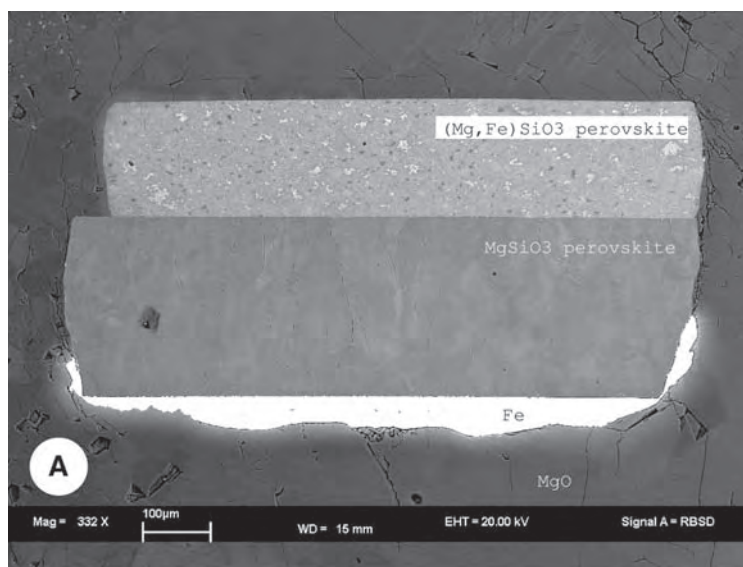
Fig. 1. (**A**) Backscattered electron image of a perovskite diffusion couple (sample C13) performed in an MgO-Fe capsule. The experiment was annealed at 24 GPa and 2043 K for 240 min. The Fe-bearing perovskite contains small inclusions of ferropericlase (light phase) and stishovite (dark phase). Measured diffusion profiles are not affected by these phases. (**B**) Fe profile measured by energy-dispersive x-ray scanning TEM analysis on sample C19 run at 24 GPa and 2023 K for 483 min (Table 1), normalized to the initial endmember compositions. The line is a fitted diffusion profile with $\log(D)$ m^2 s^{-1} = −19.4. The error bars are displayed for the statistical 3σ error calculated from peak intensities.

Table 1. Results of the perovskite diffusion experiments. X_{FeSiO_3} is the initial composition of the Fe-bearing endmember of the diffusion couple; in the case of the Mg endmember, no Fe was present at the beginning of the diffusion experiment. Time denotes experimental duration. P, pressure; No. FP, number of fitted diffusion profiles.

P (GPa)	T (K)	Sample	Time (min)	Capsule	X_{FeSiO_3}	No. FP	Log(D_{Fe-Mg}) (m^2 s^{-1})
26	1973*	C76	570	Ni-NiO	0.11	3	−18.2 ± 0.1
24	2023	C19	483	MgO-Fe	0.02–0.05	3	−19.1 ± 0.5
24	2043	C13	240	MgO-Fe	0.06	2	−18.8 ± 0.6
24	2123	C23	720	MgO-Fe	0.02–0.03	1	−18.4 ± 0.2
24	2133	C28†	1430	MgO-Fe	0.06	1	−17.6 ± 0.1
26	2173	C70	60	Ni-NiO	0.09	1	−17.0 ± 0.1
22	2273	C30	373	MgO-Fe	0.05–0.07	2	−18.2 ± 0.2

*Thermocouple failed; the temperature was estimated from the heating power with an uncertainty of ±50 K.   †Because the metal in the oxygen buffer in this experiment was largely exhausted, we did not include this data point when fitting the results.

Here, we report Fe-Mg diffusion coefficients for (Mg,Fe)SiO$_3$ perovskite determined experimentally at conditions of the uppermost part of Earth's lower mantle. High-pressure interdiffusion experiments were performed on polycrystalline perovskite diffusion couples, each consisting of a disk of initially Fe-free perovskite in contact with a second disk of Fe-bearing perovskite in which X_{Fe} = Fe/(Fe + Mg) was varied between 0.02 and 0.11 (Table 1). The perovskite samples were synthesized previously from synthetic Fe-free and Fe-bearing enstatites and from a sample of natural single crystal bronzite (6) (Table 1). Perovskite grain sizes varied between 2 and 100 μm; the Fe-bearing samples had the smaller grain sizes.

Diffusion experiments were performed at pressures of 22 to 26 GPa and temperatures between 1973 and 2273 K for up to 24 hours with a multianvil apparatus (6). The diffusion couples were contained either in Ni foil capsules with added NiO powder below the sample or in single-crystal MgO capsules with Fe foil located below the sample (Fig. 1A). Oxygen fugacities were estimated to be about 3 log-units below the iron-wüstite buffer (IW−3) for MgO capsules and at the Ni-NiO buffer (~IW+3 at 25 GPa) for Ni capsules (6). Recovered samples were sectioned perpendicular to the diffusion interface and were examined by Raman spectroscopy, electron microprobe, and transmission electron microscopy (TEM).

The experimentally produced diffusion profiles were extremely short in all samples (0.15 to 1.5 μm), even after times of up to 24 hours at temperatures in excess of 2100 K. Consequently, the interdiffusion profiles could only be measured accurately with a TEM equipped with an energy dispersive x-ray analysis system (6). Electron energy loss spectroscopy was also performed using the TEM in order

to determine $Fe^{3+}/\Sigma Fe$ ratios, which were on the order of 18 to 20% for Fe-bearing samples (6).

In addition to concentration profiles resulting from lattice diffusion, TEM observations revealed the occurrence of some diffusion along grain boundaries and subgrain boundaries. However, because lattice diffusion profiles were much shorter than the average grain size, it was possible to determine coefficients of lattice diffusion at grain-grain contacts that were unaffected by grain boundary diffusion. Diffusion profiles were symmetric with no detectable dependence on composition within the compositional range investigated (Fig. 1B). In addition, a dependence on crystallographic orientation was not detected. We used a solution of the equation describing diffusion in a semi-infinite medium (7) to fit all diffusion profiles for a constant (i.e., composition- and orientation-independent) diffusion coefficient. The error in the diffusion coefficient varies between 0.1 and 0.6 log-units (Table 1 and Fig. 2A). The consistency of the results on an Arrhenius diagram (Fig. 2A) also implies that diffusion coefficients are independent of run duration (60 to 720 min). Only one experiment (C28, Table 1) showed a slightly faster diffusivity (Fig. 2A), which is attributed to an exhaustion of the metal in the oxygen buffer. Hence, the result of this experiment was not used when fitting the data. Earlier experiments on ferropericlase, which exhibits much faster diffusion than perovskite, indicated that diffusion during heating to the final run temperature of the experiment has no effect on the derived diffusion coefficients, even for short experimental durations of a few minutes (3).

Fe-Mg interdiffusion coefficients determined with the use of Ni capsules were found

[1]Bayerisches Geoinstitut, University of Bayreuth, D-95440 Bayreuth, Germany. [2]Friedrich-Schiller-Universität Jena, Institut für Geowissenschaften, Burgweg 11, D-07749, Germany.

*Present address: Universität des Saarlandes, Institut für Werkstoffwissenschaft, Funktionswerkstoffe, Postfach 15 11 50, D-66041 Saarbrücken, Germany.
†To whom correspondence should be addressed. E-mail: c.holzapfel@mx.uni-saarland.de

Fig. 2. (A) Fe-Mg interdiffusion coefficients as a function of inverse temperature at oxidizing and reducing conditions. The slope of the regression lines for oxidizing conditions (Ni-NiO capsules) and for reducing conditions (MgO-Fe capsules) gives an activation energy at 24 GPa of 414 kJ mol⁻¹. The result of experiment C28, performed at 1860°C and 24 GPa, falls off the regression line because the Fe metal

was largely oxidized in this experiment so that the buffering capacity was exhausted. Therefore, we did not use this experiment when fitting the data. (B) Comparison of diffusion results with data for other mantle minerals. Data for wadsleyite at 15 GPa are from (28) and (29), for ferropericlase at 23 GPa from (3), and for olivine at 12 GPa are from (28). In addition, the self-diffusion coefficients for Si diffusion in silicate perovskite (8) and olivine (24) and the Mg tracer diffusion coefficient of garnet (18) are displayed. CCO, C-O buffer.
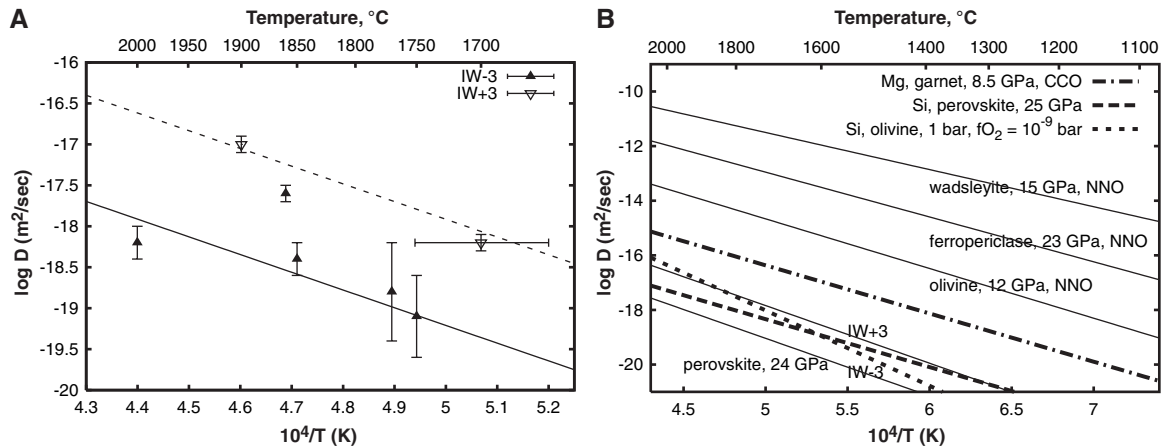
to be about one order of magnitude larger than those obtained with MgO capsules (Fig. 2A). Diffusion coefficients were therefore fitted with an Arrhenius equation, assuming that the pre-exponential factor D_0 depends on oxygen fugacity (f_{O_2}), and that the activation energy E_a is independent of f_{O_2} (Fig. 2A). This assumption implies that the change in oxygen fugacity as a function of temperature in the two capsule types is the same, which is reasonable given that the two oxygen buffers should have similar temperature dependencies. From this fit, the Fe-Mg interdiffusion coefficient D_{Fe-Mg} at 24 GPa is described by

$$D_{Fe-Mg} = D_0 \exp\left(-\frac{E_{a,24}}{RT}\right) \quad (1)$$

where D_0 has the value 4.0 ($\pm$0.7) $\times$ 10⁻⁹ m² s⁻¹ at reducing conditions ($\sim$IW–3) and 7.9 ($\pm$1.4) $\times$ 10⁻⁸ m² s⁻¹ at oxidizing conditions ($\sim$IW+3), $E_{a,24} = 414$ ($\pm$62) kJ mol⁻¹ is the activation energy at 24 GPa, R is the gas constant, and T is temperature in kelvin. The error was determined from fitting the Arrhenius equation (Eq. 1) to the diffusion coefficients given in Table 1.

The measured Fe-Mg interdiffusion coefficients are similar in magnitude to those previously determined for silicon self-diffusion in silicate perovskite (8). Under reducing conditions (MgO capsules) Fe-Mg diffusion may be even slower than silicon self-diffusion (Fig. 2B). This behavior is quite different from that exhibited by upper mantle minerals such as olivine, in which Si diffusion is several orders of magnitude slower than Fe-Mg interdiffusion (9) (Fig. 2B). This could have important implications for the rheology of the lower mantle because high-temperature diffusion creep processes are generally controlled by the slowest diffusing species, which in perovskite could be divalent cations rather than silicon. Diffusion coefficients for perovskite are more than

four orders of magnitude lower than those of ferropericlase at lower mantle conditions and are about three orders of magnitude lower than those of olivine at 12 GPa (Fig. 2B).

The variation in Fe-Mg interdiffusion coefficients for silicate perovskite can be estimated throughout the entire lower mantle if the activation volume for Fe-Mg diffusion in perovskite can be estimated. Although the pressure range in our experiments is insufficient to determine this value, a good approximation can be made by assuming that the activation volume is similar to the theoretically derived value of 2.1 cm³ mol⁻¹ for extrinsic Mg self-diffusion that was estimated to be constant over the entire pressure range of the lower mantle (6, 10). In this case, the opposing effects of pressure and temperature along a mantle geotherm (1) result in an almost constant Fe-Mg interdiffusion coefficient in the lower mantle of ~4 × 10⁻²⁰ m² s⁻¹ at reducing conditions and ~8 × 10⁻¹⁹ m² s⁻¹ at oxidizing conditions. Along a mantle adiabat, the maximum error in D is on the order of a factor of 40 (6).

Because the lower mantle contains up to 20 volume % ferropericlase, in addition to silicate perovskite, we have to evaluate the effect of this mineral on diffusion rates. An upper bound on diffusive interaction at the core-mantle boundary has been derived assuming that ferropericlase forms an interconnected network on perovskite grain boundaries, with predicted diffusion distances of up to 100 km developing over the age of Earth, particularly if grain boundary diffusion plays an important role (2). However, both experimental observations and numerical simulations show that ferropericlase forms isolated grains in a perovskite matrix in a pyrolitic bulk composition under static or low-strain rate conditions (11, 12). In addition, for likely grain sizes in the lower mantle, the effects of grain boundary diffusion should be negligible

(13). In this case, the effective bulk diffusion coefficient is estimated to be approximately twice the diffusion coefficient for perovskite (14). Oceanic crust would transform almost entirely to perovskite in the lower mantle (15), which means that the bulk diffusion coefficient will be close to the perovskite value.

Fe-Mg diffusion coefficients are of similar magnitude to those for Si self-diffusion; however, many trace elements, with larger radii and/or higher valence states, are likely to have even lower diffusion coefficients (16). The rate at which Fe-Mg equilibrium is established for a heterogeneity of a particular size in the perovskite stability field is therefore a good proxy for the minimum time scale for the homogenization of a chemical heterogeneity in the lower mantle by diffusion. The oxygen fugacity is likely low in the lower mantle and may be buffered by metallic iron (17). If this is the case, diffusion coefficients for lower mantle perovskite are likely to be closer to the lower values obtained in MgO capsules than to those obtained in Ni capsules.

Chemical heterogeneities in Earth's mantle, such as subducted oceanic crust or regions affected by metasomatism, should be removed over time as a result of high-strain deformation (thinning and stretching) and chemical diffusion (4, 5). The time $t_{0.90}$ required for chemical diffusion to equilibrate a slablike region of thickness x at oxidizing (Ni-NiO buffer, approximately 3 log-units above Fe-FeO, IW+3) and reducing ($\sim$IW–3) conditions was calculated with an analytical expression for the appropriate diffusion equation [equation 4.17 in (7) and Fig. 3]. The region is considered to be equilibrated when the average composition of the slab reaches 90% of the final value. Times obtained are similar for the homogenization of spherical regions of diameter x. At lower mantle conditions and on experimental time scales of ~24 hours, equilibrium is only achieved on a submicrometer scale under reducing condi-

Fig. 3. Equilibration times $t_{0.90}$ as a function of the size d of a chemical heterogeneity with the use of the effective Fe-Mg interdiffusion coefficient of the lower mantle for reducing (∼IW–3) and oxidizing conditions (∼IW+3), respectively. The heterogeneity is assumed to be a slab with thickness d. Also shown are the experimental time scale, the time scale for one convection cycle in the lower mantle (convection velocity = 5 cm year^{-1}), and the age of Earth (4.5×10^9 years).
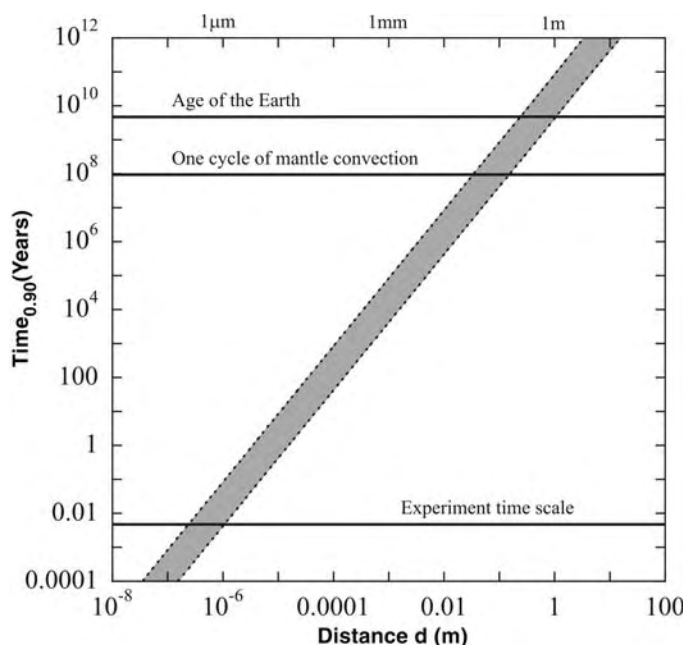
tions and over distances of a few micrometers under oxidizing conditions. In the mantle, grains with a diameter of 0.1 to 1 mm [the grain size expected for the lower mantle (12)] would equilibrate on time scales of 10 to 1000 years under oxidizing conditions or 100 to 10,000 years under reducing conditions.

During one mantle convection cycle, assuming a convection velocity of 5 cm year^{-1}, only small-scale chemical heterogeneities, with a thickness of 7 to 30 cm, are predicted to equilibrate in the lower mantle. In the upper mantle, on the basis of the likely residence time of heterogeneities and diffusion data for olivine, diffusion distances could reach ∼1 m. Although diffusion rates in wadsleyite, the stable phase in the transition zone, are much faster than in other mantle minerals (Fig. 2B), oceanic crust consists mainly of garnet in this region of the mantle (15). On the basis of rates of Mg tracer diffusion in garnet at high pressure (18), Fe-Mg interdiffusion is likely to be too slow for notable homogenization to occur in the limited depth range of the transition zone (Fig. 2B). Therefore, in the case of subducted oceanic crust, initially ∼5 km thick, negligible homogenization would occur during subduction down to the core-mantle boundary and subsequent ascent in a rising plume.

Over the entire history of Earth (4.5×10^9 years), the size of heterogeneities that could equilibrate by lattice diffusion in the lower mantle is 0.5 to 2 m depending on the redox conditions. This estimate also applies to deep regions of the lower mantle close to the D″ layer. For the thermal boundary layer at the core-mantle boundary, we calculate a diffusion distance of ∼6 m for reducing conditions and temperatures of 4500 K. Clearly, hetero-

geneities larger than a few meters can survive several cycles of convection in the lower mantle if the exchange process is controlled by lattice diffusion, which explains the observed persistence of heterogeneities in the mantle (19). Extreme deformation is required to stretch and thin heterogeneities to such an extent that equilibration distances become short enough to enable substantial chemical exchange to occur.

Recently, a phase transformation of silicate perovskite to a postperovskite phase has been inferred to occur near the D″ layer (20–22). The diffusional properties of this postperovskite phase are unknown. However, even if diffusion is faster than in silicate perovskite, the volumetric extent of the D″ layer is so small compared with the rest of the mantle that residence times of heterogeneities in this region should be short during mantle convection. Thus, the long equilibration times predicted during mantle convection (Fig. 3) are not likely to be substantially affected.

References and Notes

1. J.-P. Poirier, *Introduction to the Physics of the Earth's Interior* (Cambridge Univ. Press, Cambridge, ed. 2, 2000).
2. J. A. Van Orman, Y. Fei, E. H. Hauri, J. Wang, *Geophys. Res. Lett.* **30**, 1056 (2003).
3. C. Holzapfel, D. C. Rubie, S. Mackwell, D. J. Frost, *Phys. Earth Planet. Int.* **139**, 21 (2003).
4. P. E. Van Keken, E. H. Hauri, C. J. Bellentine, *Annu. Rev. Earth Planet. Sci.* **30**, 493 (2002).
5. P. E. Van Keken, C. J. Ballentine, E. H. Hauri, in *Treatise on Geochemistry*, H. D. Holland, K. K. Turekian, Eds. (Elsevier, Amsterdam, 2003), vol. 2, pp. 471–491.
6. Materials and methods are available as supporting material on *Science* Online.
7. J. Crank, *The Mathematics of Diffusion* (Oxford Univ. Press, New York, ed. 2, 1975).
8. D. Yamazaki, T. Kato, H. Yurimoto, E. Ohtani, M. Toriumi, *Phys. Earth Planet. Int.* **119**, 299 (2000).
9. Fe-Mg interdiffusion in olivine at 1 bar was measured

by Chakraborty (23) and Si diffusion coefficients are determined by Dohmen et al. (24).
10. For ferropericlase, it was shown that the difference in activation volume of Mg self-diffusion and Fe-Mg interdiffusion is on the order of 10% (3). We assumed that this is also true for Fe-Mg interdiffusion in perovskite and hence, the theoretically derived value of ∼2.1 cm^3 mol^{-1} (25) is a good approximation for the activation volume of perovskite.
11. N. Nishiyama, T. Irifune, T. Inoue, J. Ando, K. Funakoshi, *Phys. Earth Planet. Int.* **143–144**, 185 (2004).
12. V. S. Solomatov, R. El-Khozondar, V. Tikare, *Phys. Earth Planet. Int.* **129**, 265 (2002).
13. The governing equation to estimate the effective diffusion coefficient D_{eff} of a polycrystalline material is $D_{eff} = D_{lat} + \pi\delta D_{gbd}/g$, where D_{lat} and D_{gbd} are the lattice and grain boundary diffusion coefficients, respectively, δ denotes grain boundary width, and g is the grain size. For Si self-diffusion, the grain boundary diffusion coefficient is approximately four orders of magnitude faster than lattice diffusion (8). For Fe-Mg interdiffusion no experimental data on grain boundary diffusion exist, but a conservative estimate for the upper limit is that D_{gbd} is five orders of magnitude faster than D_{lat} and δ is about 10^{-9} m. In this case, the effective diffusion coefficient would be approximately a factor of 30 larger than the D_{lat} for a grain size of 10 μm. Equilibration times would show an enhancement on the same order, whereas diffusion length scales would be a factor of 6 longer. For grain sizes larger than 100 μm, the effect would become insignificant. The grain size of the lower mantle is estimated to be on the order of 100 to 1000 μm (12). Thus, if no substantial grain size reduction occurs, lattice diffusion should be the predominant mode of diffusion in the lower mantle.
14. Different models used were the lower Hashin-Shtrikman bound (2, 26), a series parallel formulation for composite materials (7) and numerical finite difference simulations assuming simplified geometries (27). The results from all approaches are essentially the same with an effective diffusion coefficient of $D_{eff} = 2 D_{pvsk}$ (where D_{pvsk} is the Fe-Mg interdiffusion coefficient of perovskite) for a matrix of silicate perovskite with dispersed ferropericlase inclusions.
15. S. Ono, E. Ito, T. Katsura, *Earth Planet. Sci. Lett.* **190**, 57 (2001).
16. J. A. Van Orman, T. L. Grove, N. Shimizu, *Contrib. Mineral. Petrol.* **141**, 687 (2001).
17. D. J. Frost et al., *Nature* **428**, 409 (2004).
18. S. Chakraborty, D. C. Rubie, *Contrib. Mineral. Petrol.* **122**, 406 (1996).
19. A. W. Hofmann, *Nature* **385**, 219 (1997).
20. M. Murakami, K. Hirose, K. Kawamura, N. Sata, Y. Ohishi, *Science* **304**, 855 (2004).
21. A. R. Oganov, S. Ono, *Nature* **430**, 445 (2004).
22. S.-H. Shim, T. S. Duffy, R. Jeanloz, G. Shen, *Geophys. Res. Lett.* **31**, L10603, 10.1029/2004Gl019639 (2004).
23. S. Chakraborty, *J. Geophys. Res.* **102**, 12317 (1997).
24. R. Dohmen, S. Chakraborty, H.-W. Becker, *Geophys. Res. Let.* **29**, 10.1029/2002GL015480 (2002).
25. K. Wright, G. D. Price, *J. Geophys. Res.* **98**, 22245 (1993).
26. Z. Hashin, S. J. Shtrikman, *Appl. Phys.* **33**, 3125 (1962).
27. C. Holzapfel, thesis, University of Cologne (2004).
28. C. Holzapfel, S. Chakraborty, D. C. Rubie, D. J. Frost, data not shown.
29. S. Chakraborty et al., *Science* **283**, 362 (1999).
30. We thank H. Schulze and O. Leitner for sample preparation and H. Fischer and G. Herrmannsdörfer for technical assistance. We are grateful to three anonymous referees for journal reviews.

Rising Atmospheric CO₂ Reduces Sequestration of Root-Derived Soil Carbon

James Heath,[1]* Edward Ayres,[1]† Malcolm Possell,[2]
Richard D. Bardgett,[1] Helaina I. J. Black,[3] Helen Grant,[4]
Phil Ineson,[5] Gerhard Kerstiens[1]

Forests have a key role as carbon sinks, which could potentially mitigate the continuing increase in atmospheric carbon dioxide concentration and associated climate change. We show that carbon dioxide enrichment, although causing short-term growth stimulation in a range of European tree species, also leads to an increase in soil microbial respiration and a marked decline in sequestration of root-derived carbon in the soil. These findings indicate that, should similar processes operate in forest ecosystems, the size of the annual terrestrial carbon sink may be substantially reduced, resulting in a positive feedback on the rate of increase in atmospheric carbon dioxide concentration.

The concentration of CO_2 in the atmosphere has risen from its preindustrial level of ~280 $\mu mol\ mol^{-1}$ to 376 $\mu mol\ mol^{-1}$ in the year 2003 (*1*) as a result of the combustion of fossil fuels and land-use changes such as deforestation. This continuing increase would be far more rapid were it not for the removal of large amounts of CO_2 from the atmosphere and its storage in ocean and terrestrial ecosystems; the size of the global terrestrial carbon sink during the 1990s was 2.8 ± 0.9 Gt C yr^{-1}, of which approximately one-quarter was absorbed by northern temperate and boreal forests (*2*). Because forest carbon sinks are such an important control on atmospheric CO_2 concentration and associated climate change, it is important to quantify their potential for increased carbon storage as CO_2 concentrations continue to rise, and to understand the mechanisms that determine their magnitude.

The majority of carbon stored in global vegetation is in forests. The growth of trees and the preservation of old forests are therefore of prime importance in regulating the size of the overall terrestrial carbon sink (*3*). In temperate and boreal forests, the amount of carbon stored in the soil is about four times as high as that stored in the vegetation,

and 33% higher than total carbon storage in tropical forests (*4*). Thus, there is considerable potential for long-term sequestration of carbon in the soils of temperate/boreal forests where, in biochemically stable or mineral-bound form, it could have a residence time of hundreds to thousands of years (*5*). Many studies have concentrated on the effects of elevated CO_2 on plant productivity, but the belowground exchange of carbon between plants and soil remains poorly understood (*6*), particularly in relation to the transfer of newly fixed carbon to, and storage in, long-term soil carbon pools (*7*). Fine root production and turnover represent a substantial proportion of annual net primary productivity in trees and forests (*8–10*). Our experiment was therefore designed to quantify sequestration of root-derived carbon in soil beneath a range of tree species native to much of Europe, grown at CO_2 concentrations ranging from current ambient to 300 $\mu mol\ mol^{-1}$ above ambient.

We used three pairs of tree species of contrasting shade tolerance/successional status (*11*) to represent a wide range of taxonomic, physiological, and ecological types. Trees were grown for 2 years under four CO_2 concentrations (ambient and ambient +100, +200, and +300 $\mu mol\ mol^{-1}$) to obtain response curves over this range and at two levels of soil nutrient availability to enable testing of the extent to which responses to elevated CO_2 are nutrient limited (*12–14*). To overcome the difficulties in directly measuring changes in carbon content of native forest soils (*15*, *16*), soil carbon sequestration was quantified by using stable isotope natural abundance techniques, which enabled accurate measurement of sequestration of plant-derived carbon transferred to the soil over the course of the experiment (*7*, *11*, *17*, *18*). Because leaf litter was removed from the soil (*11*), carbon inputs

over the course of the experiment were derived solely from root turnover and exudation. Trees were grown in large, deep mesocosms containing C_4 grassland soil inoculated with microbes (including mycorrhizal fungi) and mesofauna from native UK woodlands and placed in well-ventilated hemispherical greenhouses (*11*).

All tree species responded similarly to elevated CO_2 (tables S1 to S6 and figs. S5 to S11) except for a significant increase in specific leaf area of the first-season needles of *Pinus* at elevated CO_2, which caused a corresponding increase in the leaf area ratio of *Pinus* at the end of the first growing season (table S3) and a significant interaction between species and CO_2 concentration regarding the ratio of fine roots to structural roots (table S3). Therefore, the values shown in Figs. 1 to 3 are means of all six species, expressed as percentages of their low-nutrient, ambient CO_2 controls. In this way, the large absolute differences between species (tables S1 to S6) are excluded to show the overall effect of CO_2 concentration (which was the same across all species) and nutrient supply. In all species, increasing CO_2 concentration caused a decline in stomatal conductance (g_s) (Fig. 1A and table S2) and in maximum (i.e., light and CO_2 saturated) photosynthetic rate (A_{max}) (Fig. 1B and tables S1 and S2) at both soil nutrient levels. There were no changes in the specific leaf area of second-season leaves (table S3) to account for the reductions in A_{max} (which was measured on a leaf area basis). Therefore, the reduction in A_{max} was due directly to physiological adjustments rather than changes in leaf morphology. There was a 15% decrease in leaf nitrogen content from ambient to ambient +300 $\mu mol\ mol^{-1}$ CO_2 (table S4), which probably contributed to the decline in A_{max}. Despite reduced g_s and A_{max}, net photosynthetic rate (i.e., actual CO_2 uptake by the trees) was substantially increased under elevated CO_2 at both soil nutrient levels (Fig. 1C and table S2). Increased photosynthesis was associated with an initial stimulation of growth, which was only maintained where nutrients had been added (Fig. 1D and table S3). The growth response to CO_2 concentration was not linear but occurred mainly between ambient and ambient +100 $\mu mol\ mol^{-1}$.

Despite enhanced tree growth, we found a marked decline in sequestration of root-derived carbon in the soil as CO_2 concentration increased, particularly between ambient +100 and +200 $\mu mol\ mol^{-1}$ (Fig. 2 and table S5). After 15 months, soil carbon sequestration was reduced by more than 40% at the highest CO_2 concentration, relative to ambient. The addition of nutrients caused a slight increase in the amount of root-derived carbon sequestered in the soil (Fig. 2); however, the response to CO_2

[1]Department of Biological Sciences, [2]Department of Environmental Science, Institute of Environmental and Natural Sciences, Lancaster University, Bailrigg, Lancaster, LA1 4YQ, UK. [3]Soil Ecology Group, [4]Natural Environment Research Council Life Sciences Mass Spectrometry Facilities, Centre for Ecology and Hydrology, Lancaster Environment Centre, Bailrigg, Lancaster, LA1 4AP, UK. [5]Stockholm Environment Institute, Biology Department, University of York, Post Office Box 373, York, YO10 5YW, UK.

*To whom correspondence should be addressed. E-mail: j.heath@lancaster.ac.uk
†Present address: Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, CO 80523–1499, USA.

concentration was unaffected, even though plant (*12*) and soil (*19*) nitrogen status are expected to influence the effects of elevated CO_2 on soil organic matter decomposition. The reduced sequestration of root-derived carbon in the soil under elevated CO_2 was associated with an increase in soil microbial respiration as measured at the end of the experiment (*11*); the increase in carbon sequestration with added nutrients was associated with lower respiration rates (Fig. 3 and table S6). Even though irrigation rates were adjusted for a slight excess supply of water (*11*), there was a slight increase in soil moisture content at elevated CO_2 concentrations (fig. S10), which was probably due to the reduction in g_s. However, there was little loss of carbon in drainage water, and the amount was unaffected by CO_2 concentration (fig. S11). Therefore, it seems likely that reduced carbon sequestration in the soil resulted, at least in part, from enhanced turnover of root-derived carbon associated with increased microbial respiration over the course of the experiment. The possibility that respiration of the roots themselves (not measured) increased under elevated CO_2 cannot be discounted, but the existing literature does not point in this direction (*20*).

Differences in the net carbon exchange of European forests tend to be driven by changes in respiration rates at more northerly latitudes (*16*), with ~50% of total ecosystem respiration heterotrophic in nature (*21*). Therefore, any effect of increased atmospheric CO_2 concentration resulting in increased heterotrophic respiration could have important implications for the carbon balance of these forests. Increased soil moisture throughout the year may result in higher rates of respiration (*16*, *21*). However, although there was a small increase in soil moisture at elevated CO_2 concentrations in our experiment (fig. S10), this was not correlated with respiration rates ($P = 0.784$). Therefore, the increase in microbial respiration was probably a result of changes in the quantity (*6*, *15*, *19*, *22*) and quality (*6*, *12*, *19*, *23*) of inputs of root-derived organic matter from the trees.

Carbon flux to the soil may be increased by enhanced fine root production (*9*, *15*), turnover rates (*8*) or increased root exudation (*6*). Root biomass responded to CO_2 concentration and nutrient addition in a similar manner to total biomass (table S3). Within individual species, there were no significant relationships between microbial respiration and fine root biomass ($P > 0.999$). However, although we only measured root biomass at the beginning and end of the growing seasons, elevated CO_2 may have affected patterns of root mortality and regrowth over the course of the winter, resulting in the clear effect on carbon sequestration after 10 months (Fig. 2). Elevated CO_2 has also been shown to cause an increase in midseason fine root production, followed by a corresponding increase in mortality (*9*). Previous studies of trees have shown that increased fine root production and mortality under elevated CO_2 were associated with enhanced microbial activity and loss of CO_2 from soil respiration (*15*, *22*, *23*). These effects were found to be associated with changes in the composition of the microbial community, but as in our experiment (figs. S8 and S9), total microbial biomass was unaffected (*23*). An increase in respiration with no change in microbial biomass may also have been associated with faster rates of microbial turnover, for example, as a consequence of increased grazing by soil fauna (*24*).

Alternatively, the stimulation of microbial respiration may have been associated with changes in the overall or temporal quantity and quality of root exudates. Fine root carbon/nitrogen ratio at the end of the experiment was unaffected by CO_2 concentration (table S4),
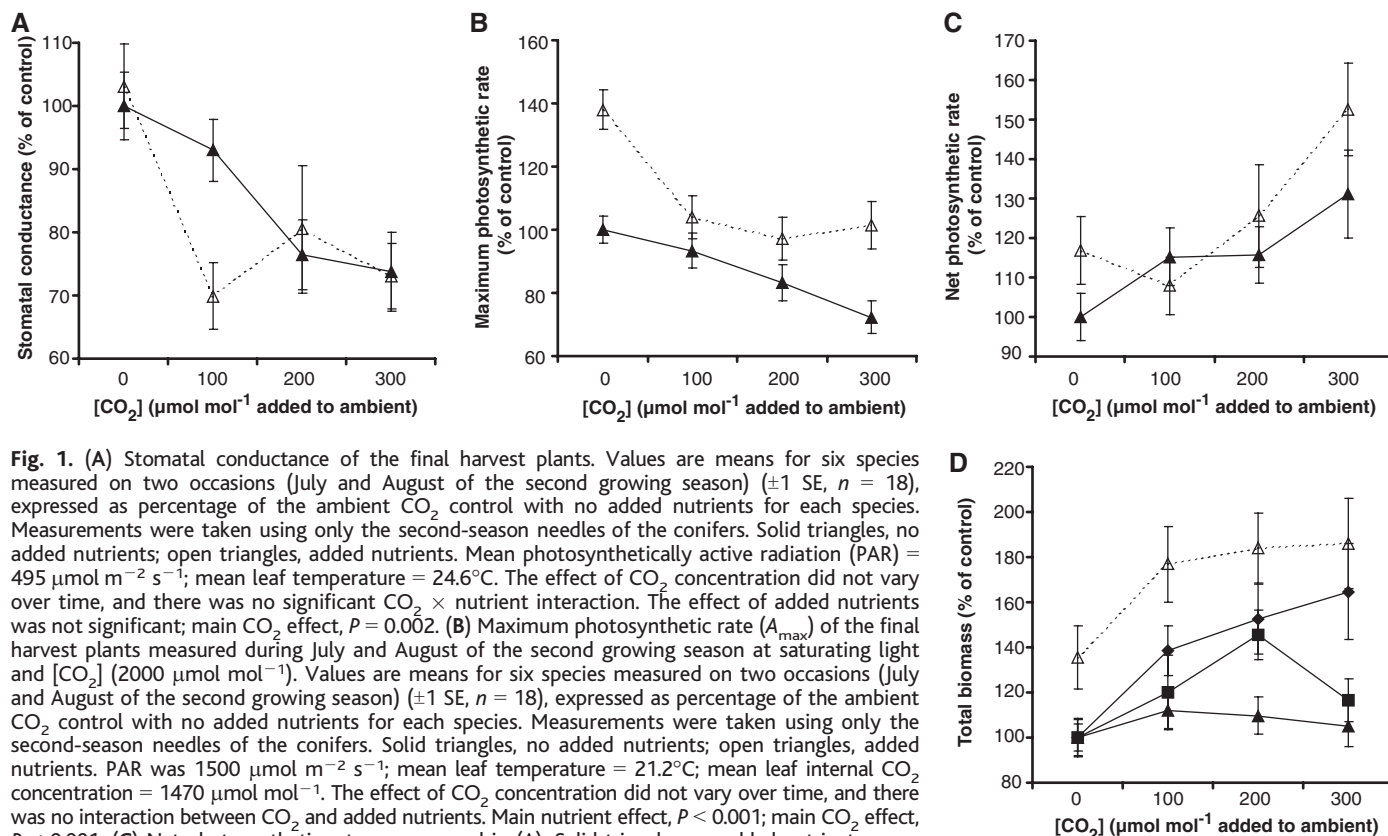
Fig. 1. (**A**) Stomatal conductance of the final harvest plants. Values are means for six species measured on two occasions (July and August of the second growing season) (± 1 SE, $n = 18$), expressed as percentage of the ambient CO_2 control with no added nutrients for each species. Measurements were taken using only the second-season needles of the conifers. Solid triangles, no added nutrients; open triangles, added nutrients. Mean photosynthetically active radiation (PAR) = 495 μmol m^{-2} s^{-1}; mean leaf temperature = 24.6°C. The effect of CO_2 concentration did not vary over time, and there was no significant $CO_2 \times$ nutrient interaction. The effect of added nutrients was not significant; main CO_2 effect, $P = 0.002$. (**B**) Maximum photosynthetic rate (A_{max}) of the final harvest plants measured during July and August of the second growing season at saturating light and [CO_2] (2000 μmol mol^{-1}). Values are means for six species measured on two occasions (July and August of the second growing season) (± 1 SE, $n = 18$), expressed as percentage of the ambient CO_2 control with no added nutrients for each species. Measurements were taken using only the second-season needles of the conifers. Solid triangles, no added nutrients; open triangles, added nutrients. PAR was 1500 μmol m^{-2} s^{-1}; mean leaf temperature = 21.2°C; mean leaf internal CO_2 concentration = 1470 μmol mol^{-1}. The effect of CO_2 concentration did not vary over time, and there was no interaction between CO_2 and added nutrients. Main nutrient effect, $P < 0.001$; main CO_2 effect, $P < 0.001$. (**C**) Net photosynthetic rate as measured in (A). Solid triangles, no added nutrients; open triangles, added nutrients. The effect of CO_2 concentration did not vary over time, and there was no significant $CO_2 \times$ nutrient interaction. The effect of added nutrients was not significant; main CO_2 effect, $P = 0.004$. (**D**) Total biomass of plants harvested after 5, 10, and 15 months. Values are means for six species (± 1 SE, $n = 18$), expressed as percentage of the ambient CO_2 control with no added nutrients for each species. Diamonds, 5-month; squares, 10-month; and solid triangles, 15-month harvests, no added nutrients; open triangles, 15-month harvest, added nutrients. There was no significant $CO_2 \times$ nutrient interaction, and the $CO_2 \times$ harvest date interaction was not significant ($P = 0.052$). Main nutrient effect, $P < 0.001$; main CO_2 effect, $P = 0.002$.

and there was no change in soil nitrogen mineralization rate (fig. S7). This suggests that the increase in respiration was associated mainly with enhanced turnover of root-derived sugars from exudation. Plants are able to actively up-regulate the quantity or alter the quality of exudates to, for example, alleviate biotic or abiotic stress or enhance nutrient acquisition (either by altering the soil chemical environment or through mutualistic associations with fungi or bacteria) (*6*). Such responses could have resulted in enhanced microbial respiration and a corresponding decline in the sequestration of root-derived carbon in the soil at elevated CO_2 concentrations.

Elevated CO_2 commonly results in enhanced plant growth and increased carbon inputs to the soil, which stimulates microbial respiration (*25*). However, although elevated CO_2 caused an increase in microbial respiration, which was associated with reduced sequestration of new carbon in mineral-bound form, in soil beneath grassland communities (*26*), no previous studies have shown a similar effect in soils growing trees. In contrast to our experiment, CO_2 enrichment caused an increase in soil carbon sequestration beneath *Betula* seedlings over the course of one growing season (*27*), but the effect on microbial respiration is not known. Free-air CO_2 enrichment (FACE) also caused an increase in the sequestration of new carbon in C_4 soil cores transplanted into former agricultural ground beneath 2- to 3-year-old *Populus* saplings; however, in one species of *Populus*, the

opposite effect was observed in the first of two growing seasons (*28*). Given the difficulties associated with directly measuring changes in soil carbon content (*15*, *16*) and in comparing carbon sequestration in native forest soil of ambient and FACE plots (*7*, *29*, *30*), there is insufficient evidence to predict with certainty whether plant responses to elevated CO_2 will result in increased or decreased sequestration of new carbon in the soils of forest ecosystems.

Our data reveal a marked decline in sequestration of root-derived carbon in the soil at elevated CO_2 concentrations in a wide range of tree species. This effect occurred independently of plant and soil nutrient status. Two caveats need to be noted. First, young trees, grown in mesocosms in a semicontrolled environment and protected from major herbivores, may respond differently from mature trees growing in a natural forest. Second, the experiment ran for only two growing seasons and the input of leaf litter to the soil was excluded. Therefore, the possibility that longer term increased inputs of leaf litter under elevated CO_2 could counteract the effect on the sequestration of root-derived carbon cannot be ruled out. Furthermore, although soil microbial respiration increased under elevated CO_2, the effect of this on the decomposition of native soil carbon is not known. Nevertheless, this study clearly demonstrates that a mechanism exists that may drastically affect the potential for sequestration of new carbon in forest soils. Even small shifts in the carbon balance of forests could cause a large feedback on atmospheric CO_2 concentration, given that the annual exchange of CO_2 in the form of terrestrial photosynthesis and respiration is approximately 9 to 10 times as large as annual emissions from the burning of fossil fuels (*3*, *21*). Our results suggest that the incorpo-

ration of root-derived carbon into stable, medium- or long-term forest soil carbon pools may be substantially reduced as atmospheric CO_2 concentration exceeds 100 $\mu mol\ mol^{-1}$ above current ambient. This would have the potential to trigger a large positive feedback on the rate of increase in global atmospheric CO_2 concentration and associated climate change.

References and Notes

1. C. D. Keeling, T. P. Whorf, in *Trends: A Compendium of Data on Global Change* (Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, TN, USA, 2004), http://cdiac.esd.ornl.gov/trends/co2/sio-mlo.htm
2. Y. Malhi, *Philos. Trans. R. Soc. London Ser. A* **360**, 2925 (2002).
3. C. Körner, *Ecol. Appl.* **10**, 1590 (2000).
4. IPCC, *Land Use, Land Use Change, and Forestry: IPCC Special Report*, R. T. Watson *et al.*, Eds. (Cambridge Univ. Press, Cambridge, 2000).
5. E. S. Krull, J. A. Baldock, J. O. Skjemstad, *Funct. Plant Biol.* **30**, 207 (2003).
6. D. L. Jones, A. Hodge, Y. Kuzyakov, *New Phytol.* **163**, 459 (2004).
7. D. E. Pataki *et al.*, *Bioscience* **53**, 805 (2003).
8. R. Matamala, M. A. Gonzàlez-Meler, J. D. Jastrow, R. J. Norby, W. H. Schlesinger, *Science* **302**, 1385 (2003).
9. R. J. Norby, J. Ledford, C. D. Reilly, N. E. Miller, E. G. O'Neill, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 9689 (2004).
10. R. M. Rytter, *For. Ecol. Manage.* **140**, 177 (2001).
11. Materials and methods are available as supporting material on *Science* Online.
12. B. A. Hungate, J. S. Dukes, M. R. Shaw, Y. Luo, C. B. Field, *Science* **302**, 1512 (2003).
13. Y. Luo *et al.*, *Bioscience* **54**, 731 (2004).
14. R. Oren *et al.*, *Nature* **411**, 469 (2001).
15. J. S. King *et al.*, *Oecologia* **128**, 237 (2001).
16. R. Valentini *et al.*, *Nature* **404**, 861 (2000).
17. E. Ayres *et al.*, *Ecol. Lett.* **7**, 469 (2004).
18. P. L. Staddon, *Trends Ecol. Evol.* **19**, 148 (2004).
19. W. Cheng, D. W. Johnson, *Plant Soil* **202**, 167 (1998).
20. M. A. Gonzàlez-Meler, L. Taneva, R. J. Trueman, *Ann. Bot. (London)* **94**, 647 (2004).
21. J. Grace, M. Rayment, *Nature* **404**, 819 (2000).
22. J. L. Larson, D. R. Zak, R. L. Sinsabaugh, *Soil Sci. Soc. Am. J.* **66**, 1848 (2002).
23. R. L. Phillips, D. R. Zak, W. E. Holmes, D. C. White, *Oecologia* **131**, 236 (2002).
24. T. H. Jones *et al.*, *Science* **280**, 441 (1998).
25. D. R. Zak, K. S. Pregitzer, J. S. King, W. E. Holmes, *New Phytol.* **147**, 201 (2000).
26. Z. G. Cardon *et al.*, *Soil Biol. Biochem.* **33**, 365 (2001).
27. P. Ineson, M. F. Cotrufo, R. Bol, D. D. Harkness, H. Blum, *Plant Soil* **187**, 345 (1996).
28. M. R. Hoosbeek *et al.*, *Global Biogeochem. Cycles* **18**, GB1040 (2004).
29. W. M. Loya, K. S. Pregitzer, N. J. Karberg, J. S. King, C. P. Giardina, *Nature* **425**, 705 (2003).
30. W. H. Schlesinger, J. Lichter, *Nature* **411**, 466 (2001).
31. We are grateful to the Natural Environment Research Council (NERC), who provided funding for this work (grant GR3/12886); stable isotope analyses were carried out at the NERC Life Sciences Mass Spectrometry Facilities (Centre for Ecology and Hydrology Lancaster node). We also thank C. Owensby and L. Auen for the extraction of C_4 soil; and D. Andrew, M. Blackburn, A. Burns, S. Dutton, M. Harrison, R. Jones, A. Keates, M. Kitschke, D. Leech, P. Nott, H. Quirk, P. Segre, and P. Smith for technical support.
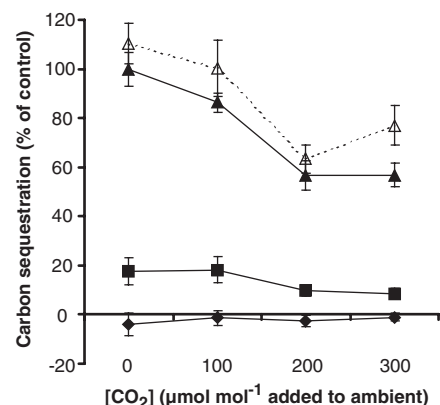
Fig. 2. Carbon sequestration in the soil after 5, 10, and 15 months. Values are means for six species (± 1 SE, $n = 18$), expressed as percentage of the ambient CO_2 control, with no added nutrients for each species at the final harvest. Absolute values of carbon sequestration at the final harvest ranged from 2.4 to 1.4 mg C g^{-1} dry soil (no added nutrients) and from 2.6 to 1.6 mg C g^{-1} dry soil (added nutrients). Diamonds, 5-month; squares, 10-month; and solid triangles, 15-month harvests, no added nutrients; open triangles, 15-month harvest, added nutrients. There was no significant $CO_2 \times$ nutrient interaction. Main nutrient effect, $P = 0.005$; main CO_2 effect, $P < 0.001$; $CO_2 \times$ harvest date interaction, $P < 0.001$.

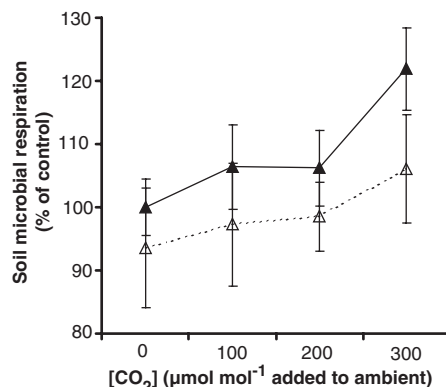**Fig. 3.** Soil microbial respiration measured after the final harvest. Values are means for six species (± 1 SE, $n = 18$), expressed as percentage of the ambient CO_2 control with no added nutrients for each species. Solid triangles, no added nutrients; open triangles, added nutrients. There was no significant $CO_2 \times$ nutrient interaction. Main nutrient effect, $P = 0.024$; main CO_2 effect, $P = 0.038$.

Unexpected Changes to the Global Methane Budget over the Past 2000 Years

D. F. Ferretti,[1,2]* J. B. Miller,[3] J. W. C. White,[1] D. M. Etheridge,[4]
K. R. Lassey,[2] D. C. Lowe,[2] C. M. MacFarling Meure,[4] M. F. Dreier,[1]
C. M. Trudinger,[4] T. D. van Ommen,[5] R. L. Langenfelds[4]

We report a 2000-year Antarctic ice-core record of stable carbon isotope measurements in atmospheric methane ($\delta^{13}CH_4$). Large $\delta^{13}CH_4$ variations indicate that the methane budget varied unexpectedly during the late preindustrial Holocene (circa 0 to 1700 A.D.). During the first thousand years (0 to 1000 A.D.), $\delta^{13}CH_4$ was at least 2 per mil enriched compared to expected values, and during the following 700 years, an about 2 per mil depletion occurred. Our modeled methane source partitioning implies that biomass burning emissions were high from 0 to 1000 A.D. but reduced by almost ~40% over the next 700 years. We suggest that both human activities and natural climate change influenced preindustrial biomass burning emissions and that these emissions have been previously understated in late preindustrial Holocene methane budget research.

Methane is an important greenhouse gas, and measurements of its atmospheric concentration, [CH_4], from ice cores demonstrate a slow increase over the late preindustrial Holocene (the LPIH, circa 0 to 1700 A.D.) (1) and a rapid increase to unprecedented levels over recent centuries (1–3). However, the causes of these variations are not well understood (1–3), because emission rates from the diverse methane sources are spatially and temporally variable, sometimes small, and distributed globally. Additionally, methane emissions vary with climate (1) and possibly with preindustrial human activities such as rice cultivation, cattle farming, and biomass burning (4, 5).

The sources of atmospheric methane can be separated into three general categories based on their stable carbon isotope ratios $\delta^{13}CH_4$: biogenic [e.g., wetlands, $\delta^{13}CH_4$ near –60 per mil (‰)], fossil ($\delta^{13}CH_4$ near –40‰), and pyrogenic or biomass burning ($\delta^{13}CH_4$ near –25‰ for C_3 vegetation or –12‰ for C_4 vegetation). Changes in atmospheric $\delta^{13}CH_4$ allow the contributions from each of these source types to be deduced (6). However, very few long $\delta^{13}CH_4$ records exist (6), because the small amount of air available from ice cores

does not generally meet the large sample size requirements for gas isotope analyses (7). Furthermore, firn-air samples are at most ~100 years old (8–10). To overcome these problems, we used a high-precision technique that was specially adapted (11) to analyze very small volume air samples extracted from Law Dome ice cores with high temporal resolution. We present 2000-year $\delta^{13}CH_4$ and [CH_4] records (Fig. 1) that reveal unexpected $\delta^{13}CH_4$ features. We used those measurements to evaluate methane sources and enhance our understanding of the LPIH methane budget. Shorter term $\delta^{13}CH_4$ variations within the industrial era and the unconfirmed role of preindustrial methane sink variations are investigated elsewhere (12).

On the basis of a greater preindustrial dominance of wetland sources and relatively stable [CH4] (Fig. 1), we expected LPIH atmospheric $\delta^{13}CH_4$ to be stable and isotopically depleted relative to the present day [in the range about –54‰ to –49‰ (13–15)]. However, our $\delta^{13}CH_4$ data do not follow those expectations (Fig. 1). Two fundamental conundrums emerge. First, LPIH $\delta^{13}CH_4$ is at least 2‰ more enriched than expected from 0 to 1000 A.D. Second, in contrast to relatively stable [CH_4], which varies by no more than ~55 parts per billion (ppb) from 1000 to 1700 A.D., the $\delta^{13}CH_4$ measurements reveal a large ~2‰ depletion.

To quantify the source evolution, we used an atmospheric box model (13) that includes global biogenic, pyrogenic, and fossil sources and accommodates the dynamics of [CH_4] and $\delta^{13}CH_4$ responses to budget changes. We adopted a weighted average of the kinetic isotope effects of methane sinks ($k_{13}/k_{12} - 1$) of –7.4‰. This sink fractionation gives consistency between [CH_4] and $\delta^{13}CH_4$ with the source construction of the EDGAR-HYDE 1.4

data set (16) over 1890 to 1995 A.D. Although OH may have been ~10% higher in the preindustrial compared to the present day (15, 17), it is likely to have been relatively much more stable from 0 to 1700 A.D., and we adopted a constant LPIH methane lifetime of 7.6 years (18). On the basis of previous studies (13–15), we postulated constant LPIH fossil methane emissions of 20 Tg•year^{-1}, because there is no evidence of elevated fossil emissions from 0 to 1000 A.D. relative to 1000 to 1700 A.D. [even though they may be higher than 20 Tg•year^{-1} (19)]. Thus, our inverse source partitioning was tightly constrained, because the total methane source strength (over time) was derived from observed [CH_4] and the partitioning of biogenic and pyrogenic sources was derived from observed $\delta^{13}CH_4$. For consistency with the relatively high levels of reported preindustrial C_4 grassland burning (4, 20, 21), we specified the C_3:C_4 plant type mix in the biomass fuel as 60:40 (i.e., a weighted-mean $\delta^{13}CH_4$ of about –20‰).

Our modeled pyrogenic and biogenic source emissions were ~25 ± 1 Tg•year^{-1} and ~194 ± 2 Tg•year^{-1}, respectively, from 0 to 1000 A.D. (Fig. 2) (22). Our preindustrial pyrogenic source is at least a factor of 2.5 larger than expected (14, 15). However, high levels of LPIH burning have been proposed before. Subak (4), using historical land-use records and current estimates of methane emission factors, estimated global pyrogenic emissions in 1500 A.D. at ~26 Tg•year^{-1}. The biogenic source is ~10% lower than the Houweling et al. (15) best estimate at 1700 A.D. (218 Tg•year^{-1}) but well within their estimated upper and lower limits (118 to 325 Tg•year^{-1}). Our source reconstruction requires that, by 1700 A.D., pyrogenic emissions decrease by ~40% to ~15 ± 1 Tg•year^{-1} (relative to emissions from 0 to 1000 A.D.) and that biogenic emissions simultaneously increase by ~10% to ~210 ± 2 Tg•year^{-1} (Fig. 2). These methane source variations may be caused by natural climate change or human activities.

Variations in temperature and moisture can influence natural methane emissions from wetlands and wildfires. If warmer temperatures coincide with dryer conditions, then during warm-dry periods we would expect elevated pyrogenic emissions and reduced biogenic emissions compared to those in cool-wet climates. During warm-dry periods, we expect that, even though temperature would increase wetland emission rates (per area), increased evaporation would reduce wetland extent, causing lower net biogenic emissions. Temperature and moisture patterns vary regionally. However, regional records of drought, rainfall, and biomass burning recovered from lakes in Africa (23), Asia (24), Europe (25), Oceania (25), and South, Central, and North America (25–27), together with chemical records of high-latitude Northern Hemispher-

[1]Institute of Arctic and Alpine Research, University of Colorado, Boulder, CO 80309, USA. [2]National Institute of Water and Atmospheric Research Limited (NIWA), Post Office Box 14901, Wellington, New Zealand. [3]National Oceanic and Atmospheric Administration (NOAA), R/CMDL1, 325 Broadway, Boulder, CO 80305, USA. [4]Commonwealth Scientific and Industrial Research Organisation (CSIRO), Marine and Atmospheric Research, Private Bag 1, Aspendale, Victoria 3195, Australia. [5]Department of the Environment and Heritage, Australian Antarctic Division, and Antarctic Climate and Ecosystems CRC, Private Bag 80, Hobart, Tasmania 7001, Australia.

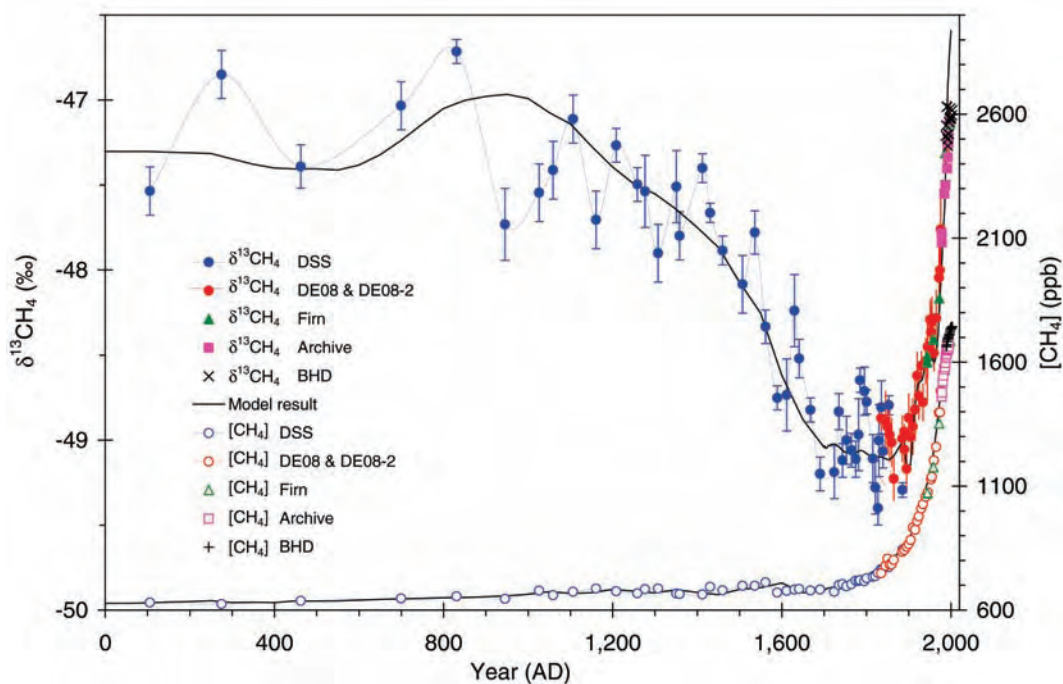*To whom correspondence should be addressed. E-mail: d.ferretti@niwa.co.nz

Fig. 1. The 2000-year Law Dome records of $\delta^{13}CH_4$ and $[CH_4]$. Air samples are from ice cores (DSS, DE08, and DE08-2), firn, archives (Cape Grim, Australia), and Baring Head, New Zealand (BHD) (*11*). The modeled result shows that the source evolution discussed in the text reproduces the main features of atmospheric $\delta^{13}CH_4$ evolution while matching $[CH_4]$. $\delta^{13}CH_4$ errors represent both measurement and diffusion correction uncertainties. $[CH_4]$ and dating errors are smaller than plotted symbols.

ic biomass burning from a Greenland ice core (*28*), provide supporting evidence that moisture is negatively correlated to temperature on a larger scale and that the extent and magnitude of methane emissions from wildfires decreased and that from wetlands increased in response to cooling temperatures and increasing moisture from ~1000 to 1700 A.D. Several of the multicentury and multidecadal anomalies in reconstructed Northern Hemispheric temperatures (*29, 30*) correlate with $\delta^{13}CH_4$ until ~1500 A.D. (Fig. 3A), providing further supporting evidence that natural climate change influenced the methane budget as we propose until ~1500 A.D.

Further evidence for the influence of pyrogenic emissions on the methane budget comes from the Law Dome carbon monoxide concentration ([CO]) record (Fig. 4). CO is a trace gas proxy for preindustrial variations in pyrogenic emissions, especially for combustion of woody biomass, which tends to be more incomplete and productive of CO than the burning of grasslands (*31*). The close correspondence between [CO] and $\delta^{13}CH_4$ from 0 to 1500 A.D. provides supporting evidence that variations in woody pyrogenic emissions were important between 0 and 1500 A.D. (*18*). However, from 1500 to 1700 A.D., [CO] remains relatively stable, whereas $\delta^{13}CH_4$ declines by another ~1‰ (Fig. 4) and annual pyrogenic methane emissions decline by another ~5 Tg (Fig. 2). A relatively larger decrease in grassland burning from 1500 to 1700 A.D. is consistent with these criteria, because, compared to woody burning, it does not produce as much CO relative to methane and it has stronger $\delta^{13}CH_4$ leverage. From ~1500 to 1700 A.D., the weakened Northern Hemisphere temperature–
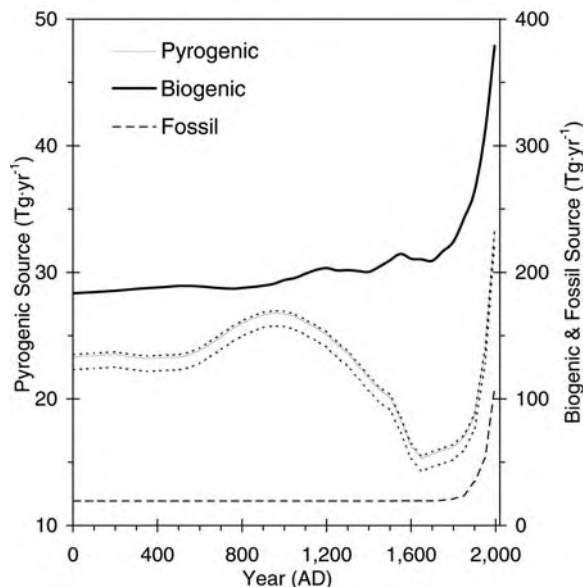

Fig. 2. Reconstructed variations in global methane source emissions. Pyrogenic source emissions are shown on the left axis; biogenic and fossil source emissions are shown on the right axis. Pyrogenic and biogenic methane source emission errors are ±1 Tg·year⁻¹ and ±2 Tg·year⁻¹, respectively (*22*). If the C_3:C_4 pyrogenic mix changed over time, such as between 70:30 and 30:70 (source $\delta^{13}CH_4$ between −21 and −16‰), the pyrogenic source emission would be constrained within the area between the thin dotted lines. Extensions of the simulation beyond 1700 A.D. are shown for comparison.

$\delta^{13}CH_4$ correlation (Fig. 3A) therefore suggests a reduced influence of climatic change on biomass burning variability during this time.

The factor most likely to have influenced pyrogenic variations from 1500 to 1700 A.D. is human activity. Estimated LPIH human population trends (*32, 33*) are shown in Fig. 3B. Because grasslands and forests in Europe and China were mostly cleared by 0 A.D. for agricultural or habitable lands (that were not burnt at large scale again) (*5*), and because relative changes in the African population were small (as compared to the Americas) (*32*), the relatively small indigenous population of the Americas would have had a disproportionate influence on LPIH anthropo-

genic methane emissions from fires. Their fires are very likely to have been important because it has been suggested that they burnt very large grassland areas annually (*4, 20, 21*) and maintained large-scale, smoldering woody fires in the Amazon to produce charcoal for improved soil fertility (*34*). On the basis of land use practices and present-day methane emission factors, independent studies have estimated pyrogenic CH_4 emissions in the Americas at 1500 A.D. to be as large as 10 Tg·year⁻¹ (*4*) and 8.25 Tg·year⁻¹ (*20*). However, the indigenous population of the Americas declined by 90% from 1500 to 1600 A.D. (*33*) because of the introduction of diseases by European explorers. Consequently, pyrogenic emissions

from the Americas must have reduced. The simultaneity from 1500 to 1600 A.D. of the rapid changes in the population of the Americas, $\delta^{13}CH_4$, and [CH$_4$] provides support for our hypothesis that rapid human population decline contributed substantially to the total LPIH biomass burning reduction and global $\delta^{13}CH_4$ depletion (perhaps by as much as a ~5 Tg reduction in annual emissions).

In the absence of dependable preindustrial values, we postulate constant natural fossil emissions (20 Tg•year^{-1}) and a constant C$_3$:C$_4$ ratio of burnt biomass (60:40). Even if the C$_3$:C$_4$ plant type mix in the biomass fuel

changed (as [CO] suggests from 1500 to 1700 A.D.), the overall shape of the inferred pyrogenic source evolution would be unchanged (Fig. 2). Thus, our linkage between LPIH global biomass burning and $\delta^{13}CH_4$ is robust, leading to our conclusion that biomass burning substantially affected the LPIH methane budget.

Our modeled methane source partitioning implies that a ~10 ± 1 Tg reduction in annual global biomass burning emissions of methane is the main cause of the ~2‰ $\delta^{13}CH_4$ depletion from 1000 to 1700 A.D. Simultaneously, a compensatory growth in biogenic methane emissions causes a small ~50 ppb [CH$_4$] in-

crease. Both natural and anthropogenic sources are likely to have contributed to the changes, such that (i) from 1000 to 1500 A.D., natural climatic change (becoming cooler and wetter) is the most likely cause for a reduced incidence of wildfires and an increased wetland area, and (ii) from 1500 to 1700 A.D., regional human population variations are the most likely causes of reduced pyrogenic emissions. Anthropogenic expansion of rice and ruminant agriculture (4, 5) may have also contributed to increasing natural wetland emissions from 0 to 1700 A.D. We therefore suggest that humans played a much larger than expected role in the evolution of the LPIH methane budget. Our work corroborates independent assessments that preindustrial anthropogenic pyrogenic emissions approximated those of today (4, 20) and therefore suggests that pyrogenic emissions have been previously understated in LPIH methane budget research (14, 15).

Fig. 3. Natural climate and human population variations. (**A**) The relationship between $\delta^{13}CH_4$ (dark blue circles) and Northern Hemisphere (NH) temperature anomaly reconstructions [relative to the 1961 to 1990 mean; Jones and Mann (29), blue line and shading (1σ error); Moberg et al. (30), green line and shading (largest errors at 95% confidence interval). Because atmospheric methane is short-lived (~10 years), the Jones and Mann record, which mainly uses tree ring data of decadal resolution and provides information on both temperature and moisture, is likely to better represent methane budget changes until ~1500 A.D., when the relationship with $\delta^{13}CH_4$ diminishes (c.f. multicentennial changes that the Moberg record better incorporates by use of low-resolution proxies). (**B**) Regional human population variations (32, 33).
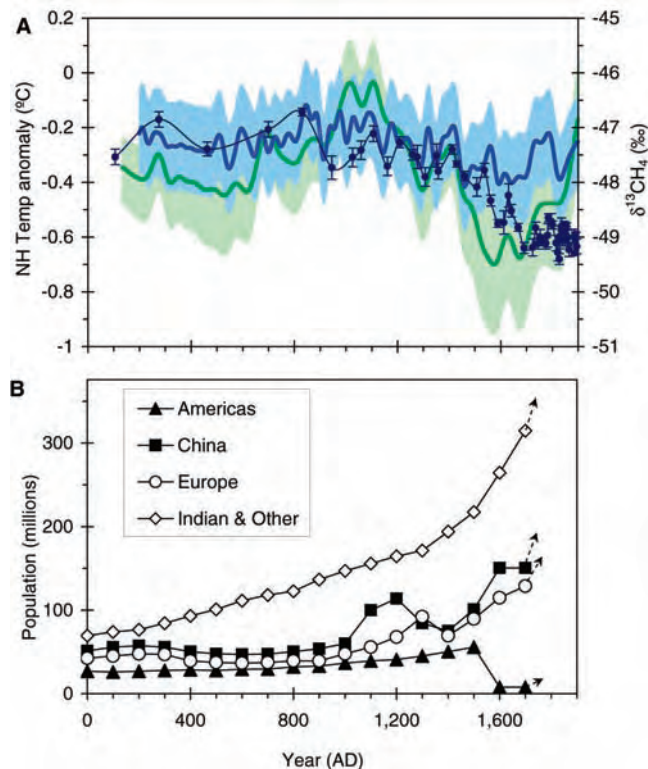
By incorporating land-use practices of different regions (i.e., biomass burning in the Americas), we investigate the relationship between population and anthropogenic source variation. Of most significance is the substantial population decline in the Americas from ~1500 A.D. "Indian & Other" refers to the remainder of the world population (which is mainly Indian).

Fig. 4. Law Dome $\delta^{13}CH_4$ (black circles) and [CO] (gray triangles) (35). CO is a short-lived trace gas proxy for biomass burning, especially of woody biomass in the preindustrial era. The [CO]–$\delta^{13}CH_4$ relationship extends over 0 to 1500 A.D. but less so during 1500 to 1700. Measurements of [CO] in the ice core have been corrected for an extraction system contamination of 6 ppb.

The error bars reflect the uncertainty in the measurement and in the storage of CO in the ice.
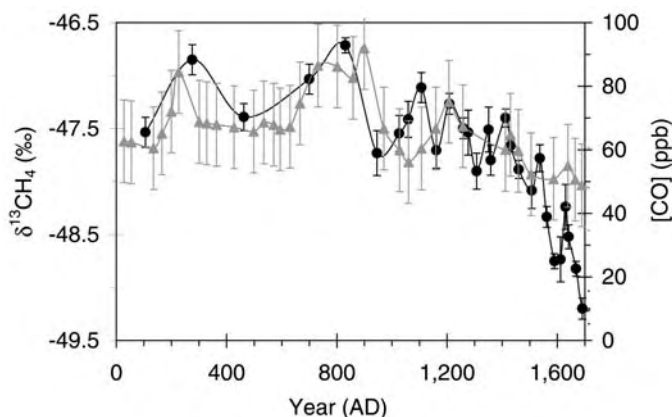
References and Notes
1. R. A. Rasmussen, M. A. K. Khalil, *J. Geophys. Res.* **89**, 11599 (1984).
2. D. M. Etheridge, L. P. Steele, R. J. Francey, R. L. Langenfelds, *J. Geophys. Res.* **103**, 15979 (1998).
3. J. R. Petit et al., *Nature* **399**, 429 (1999).
4. S. Subak, *Chemosphere* **29**, 843 (1994).
5. W. F. Ruddiman, *Clim. Change* **61**, 261 (2003).
6. H. Craig, C. C. Chou, J. A. Welhan, C. M. Stevens, A. Engelkemeir, *Science* **242**, 1535 (1988).
7. D. C. Lowe, M. R. Manning, G. W. Brailsford, A. M. Bromley, *Geophys. Res. Lett.* **24**, 857 (1997).
8. T. Sowers et al., *Global Biogeochem. Cycles* **19**, 101029/2004GB002408 (2005).
9. M. Bräunlich et al., *J. Geophys. Res.* **106**, 20465 (2001).
10. R. J. Francey et al., *J. Geophys. Res.* **104**, 23631 (1999).
11. Materials and methods are available as supporting material on *Science* Online.
12. D. Ferretti et al., in preparation.
13. K. R. Lassey, D. C. Lowe, M. R. Manning, *Global Biogeochem. Cycles* **14**, 41 (2000).
14. J. B. Miller, thesis, University of Colorado at Boulder (1999).
15. S. Houweling, F. Dentener, J. Lelieveld, *J. Geophys. Res.* **105**, 17243 (2000).
16. J. A. van Aardenne, F. J. Dentener, J. G. J. Olivier, C. G. M. Klein Goldewijk, J. Lelieveld, *Global Biogeochem. Cycles* **15**, 909 (2001).
17. Y. Wang, D. J. Jacob, *J. Geophys. Res.* **103**, 31123 (1998).
18. Even if LPIH variations in OH were as large as 10%, the weighted average fractionation of all sinks would only vary by ~0.15‰, which is equivalent to $\delta^{13}CH_4$ measurement uncertainty, and [CH$_4$] would vary by ~50 ppb, which is equivalent to observed [CH$_4$] variability. The CO variations present in Fig. 4 over 0 to 1500 A.D. could cause OH abundance to vary by up to ~10%. The 0.3°C temperature variations (Fig. 3) would cause a small change in the OH rate constant; however, $\delta^{13}CH_4$ would vary by less than 0.01 and [CH$_4$] would vary by ~4 ppb.
19. G. Etiope, *Atmos. Environ.* **38**, 3099 (2004).
20. D. W. Woodcock, P. V. Wells, *Chemosphere* **29**, 935 (1994).
21. K. Anderson, *Chemosphere* **29**, 913 (1994).
22. The modeled emission uncertainties reflect the different $\delta^{13}CH_4$ leverage of each source and result from [CH$_4$] and $\delta^{13}CH_4$ uncertainties during ice-core extraction, analysis, and calibration.
23. D. Verschuren, R. Laird, D. F. Cumming, *Nature* **403**, 410 (2000).
24. G. Ren, *Geophys. Res. Lett.* **25**, 1931 (1998).
25. C. Carcaillet et al., *Chemosphere* **49**, 845 (2002).
26. D. Hallett, R. Mathewes, R. Walker, *Holocene* **13**, 751 (2003).
27. T. W. Swetnam, *Science* **262**, 885 (1993).

28. J. Savarino, M. Legrand, *J. Geophys. Res.* **103**, 8267 (1998).
29. P. D. Jones, M. E. Mann, *Rev. Geophys.* **42**, 2003RG000143 (2004).
30. A. Moberg, D. M. Sonechkin, K. Holmgren, N. M. Datsenko, W. Karlen, *Nature* **433**, 613 (2005).
31. M. O. Andreae, P. Merlet, *Global Biogeochem. Cycles* **15**, 955 (2001).
32. C. McEvedy, R. Jones, *Atlas of World Population History* (Penguin, London, 1978), pp. 368.
33. W. Denevan, *The Native Population of the Americas in 1492* (Univ. of Wisconsin Press, Madison, WI, 1992), pp. 386.
34. B. Glaser, L. Haumaier, G. Guggenberger, W. Zech, *Naturwissenschaften* **88**, 37 (2001).
35. C. MacFarling Meure, thesis, University of Melbourne (2004).
36. We thank the staff of the Australian Antarctic Program, especially Casey Station, for field support; A. Smith for firn-air sampling assistance; the Bureau of Meteorology (Australia) for Cape Grim archive-air collection assistance; R. Francey, P. Steele, C. Allison, and S. Coram at CSIRO for logistical and technical help; and especially B. Ruddiman and B. Allan for valuable discussions. Supported by NSF (grant no. OPP0087357); NOAA/Climate Modeling and Diagnostics Laboratory; NIWA, New Zealand (Foundation for Research Science and Technology grant no. C01X0204); and the Australian Government's Antarctic Climate and Ecosystems Cooperative Research Centre and CSIRO Atmospheric Research.

Microcephalin, a Gene Regulating Brain Size, Continues to Evolve Adaptively in Humans

Patrick D. Evans,[1,2] Sandra L. Gilbert,[1] Nitzan Mekel-Bobrov,[1,2] Eric J. Vallender,[1,2] Jeffrey R. Anderson,[1] Leila M. Vaez-Azizi,[1] Sarah A. Tishkoff,[4] Richard R. Hudson,[3] Bruce T. Lahn[1]*

The gene *Microcephalin* (*MCPH1*) regulates brain size and has evolved under strong positive selection in the human evolutionary lineage. We show that one genetic variant of *Microcephalin* in modern humans, which arose ~37,000 years ago, increased in frequency too rapidly to be compatible with neutral drift. This indicates that it has spread under strong positive selection, although the exact nature of the selection is unknown. The finding that an important brain gene has continued to evolve adaptively in anatomically modern humans suggests the ongoing evolutionary plasticity of the human brain. It also makes *Microcephalin* an attractive candidate locus for studying the genetics of human variation in brain-related phenotypes.

The most distinct trait of *Homo sapiens* is the exceptional size and complexity of the brain (*1*, *2*). Several recent studies have linked specific genes to the evolution of the human brain (*3–12*). One of these is *Microcephalin* (*7*, *8*); mutations in this gene cause primary microcephaly [MCPH; Online Mendelian Inheritance in Man (OMIM) accession 251200] (*13*, *14*). MCPH is defined clinically as severe reductions in brain size coupled with mental retardation, but remarkably, an overall retention of normal brain structure and a lack of overt abnormalities outside of the nervous system (*15–17*). This led to the notion that the brains of MCPH patients function normally for their size and that genes underlying MCPH are specific developmental regulators of brain size (*15–17*).

Microcephalin is one of six known loci, named *MCPH1* through *MCPH6*, for which recessive mutations lead to MCPH (*14*, *18–23*). For four of these, the underlying genes have been identified as *Microcephalin* (*MCPH1*), *CDK5RAP2* (*MCPH3*), *ASPM* (*MCPH5*), and *CENPJ* (*MCPH6*) (*14*, *21*, *23*). Patients with loss-of-function mutations in *Microcephalin* have cranial capacities about 4 SD below the mean at birth. As adults, their typical brain size is around 400 cm³ (whereas the normal range is 1200 to 1600 cm³), and the cerebral cortex is especially small (*13*, *14*). *Microcephalin* is suggested to control the proliferation and/or differentiation of neuroblasts during neurogenesis. This postulate was consistent with several observations. First, mouse *Microcephalin* is expressed prominently in the proliferative zones of the embryonic brain (*14*). Second, the Microcephalin protein contains several copies of the BRCT domain that is found in cell cycle regulators, such as *BRCA1* (*14*, *24*). Finally, cell culture studies indeed suggested a role of *Microcephalin* in regulating cell cycle (*25–27*).

The finding that *Microcephalin* is a critical regulator of brain size spurred the hypothesis that it might have played a role in brain evolution (*16*, *28*). Consistent with this hypothesis, phylogenetic analysis of *Microcephalin* revealed signatures of strong positive selection in the lineage leading to humans (*7*, *8*). Here, we examine the possibility that positive selection has continued to operate on this gene after the emergence of anatomically modern humans.

The human *Microcephalin* locus has 14 exons spanning about 236 kb on chromosome 8p23 (*14*) (Fig. 1). We previously sequenced all the exons in 27 humans (*8*). When reanalyzing the data, we noticed that one haplotype had a much higher frequency than the other haplotypes. Additionally, this haplotype differed consistently from the others at position 37995 of the genomic sequence (counting from the start codon) or position 940 of the open reading frame. This polymorphism falls in exon 8 and changes amino acid residue 314 from an ancestral aspartate to a histidine. (This polymorphism is described as G37995C with G denoting the ancestral allele.)

To investigate whether positive selection has acted on the high-frequency haplotype, we resequenced 23.4 kb of a 29-kb region centered around the G37995C polymorphism (Fig. 1). Sequencing was performed on a panel of 89 individuals from the Coriell Institute, which broadly represents human diversity (see SOM). To assign the ancestral state of polymorphisms, we also sequenced the common chimpanzee. Several GC-rich segments were not sequenced because of technical difficulties. The resulting sequence data contained 220 polymorphic sites, including 213 single-nucleotide polymorphisms (SNPs) and 7 insertion/deletion polymorphisms (indels) (table S1).

Haplotypes were inferred using the PHASE 2.1 program (*29*, *30*). A total of 86 haplotypes

[1]Howard Hughes Medical Institute, Department of Human Genetics, [2]Committee on Genetics, [3]Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637, USA. [4]Department of Biology, University of Maryland, College Park, MD 20742, USA.

*To whom correspondence should be addressed. E-mail: blahn@bsd.uchicago.edu

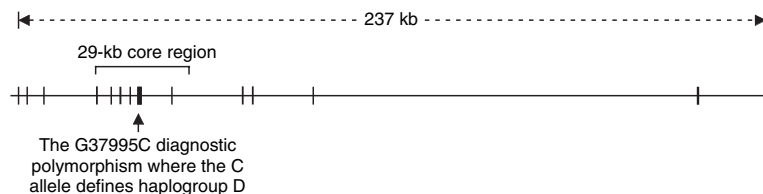**Fig. 1.** Genomic structure of the human *Microcephalin* gene. The region sequenced in the 89-individual Coriell panel is bracketed.

were identified along with their frequencies (Fig. 2 and table S2). One haplotype, denoted 49, had a much higher frequency than the other haplotypes. It had the derived C allele at the G37995C SNP site and corresponded to the high-frequency haplotype in the aforementioned exon-only polymorphism survey (8). In the Coriell panel, haplotype 49 had a frequency of 33% (59 out of 178 chromosomes) and is found in all the populations sampled in the panel. The remaining 85 haplotypes varied in frequency from 0.6 to 6.2% (1 to 11 chromosomes).

Positive selection on an allele can increase the frequency of the haplotype bearing the allele while maintaining extended linkage disequilibrium (LD) around that allele (31–36). Our data on haplotype 49 are consistent with these signatures of selection. We formally tested the statistical significance of positive selection using the previously established coalescent model (37, 38). Given the slight uncertainty in haplotype inference, we considered only the 18 individuals in the Coriell panel who are homozygous for haplotype 49 (table S1).

By simulation, we calculated the probability of obtaining 18 or more individuals (out of 89) who are homozygous for a single haplotype across a region of 220 segregating sites under neutral evolution. Here, recombination and gene conversion rates were set to values previously established for the *Microcephalin* locus (39), and a demographic model with a severe bottleneck followed by exponential growth was assumed (see SOM). Prior studies have shown that the bottleneck specified here is likely to be much more stringent than that associated with the real demographic history of human populations (40, 41); thus, the test is conservative (38). Under these parameters, the probability of obtaining 18 homozygotes out of 89 is highly significant ($P = 0$ based on 5,000,000 replicates).

We then tested several additional demographic models, including (i) constant size, (ii) very ancient expansion, (iii) very recent expansion, (iv) repeated severe bottlenecks with subsequent expansion, and (v) population structure with between two and five subpopulations (see SOM). All produced exceedingly significant results. Even though the exact demographic history of humans is yet to be defined, our tests are highly significant under a broad range of demographic scenarios, which furthers the argument that the statistical significance is unlikely to be altered by reasonable variations in the supposed human demography. We also tested the significance of the inferred haplotype data (i.e., the significance of having 59 copies of haplotype 49 among 178 chromosomes), which similarly produced highly significant results. These data strongly suggest that haplotype 49 was driven to high frequency by positive selection. However, our data do not address whether the positive selection is

frequency-dependent selection, heterozygote advantage, or simple additive positive selection.

Using the G37995C polymorphism as a diagnostic site, we divided all the haplotypes into two groups: those that carry the derived C allele and those that carry the ancestral G allele. We designated the former group as haplogroup D (where D stands for "derived"). It includes 43 haplotypes that together have a 70% frequency in the Coriell panel, and haplotype 49 is the predominant member (table S2). Although the derived C allele at the G37995C site only provides an operational definition for haplogroup D, several observations make evident that haplogroup D is systematically different from the non-D haplotypes. First, this haplogroup consists exclusively of haplotype 49 or its minor variants, whereas non-D haplotypes show much greater sequence divergence from haplogroup D chromosomes. This greater divergence is because haplogroup D and non-D haplotypes have multiple fixed differences relative to each other in addition to G37995C (table S2). The only exceptions are a few recombinant haplotypes between D and non-D chromosomes (discussed below). Second, for sites that are polymorphic within haplogroup D chromosomes (excluding recombinants between D and non-D chromosomes), the non-D chromosomes are invariably monomorphic for the ancestral alleles. These data indicate that haplogroup D constitutes a genealogical clade of closely related haplotypes that is altogether separate from the more distantly related non-D haplotypes (again, excluding recombinants between D and non-D chromosomes, which represent mixed genealogies).

Collectively, the above observations support an evolutionary scenario with two aspects.

First, haplotype 49 swept from a single copy to high frequency in a short period of time. Second, during the sweep, minor variants of haplotype 49 emerged through rare mutations and recombinations. These variants, together with haplotype 49, make up haplogroup D. Haplotype 49 evidently represents the most recent common ancestor (MRCA) of haplogroup D, because it consistently has the ancestral allele for the sites polymorphic within haplogroup D.

We next estimated the coalescence age (i.e., time to MRCA) of haplogroup D chromosomes in the Coriell panel. We used the average number of mutations from the MRCA of a haplogroup clade to its descendant lineages as a molecular clock for estimating the age of the clade (42, 43). This approach is known to be unbiased by demographic history (42). The age of haplogroup D was found to be ~37,000 years, with a 95% confidence interval of 14,000 to 60,000 years. In comparison, the coalescence age of all the chromosomes in the Coriell panel is about 1,700,000 years. The emergence of anatomically modern humans has been estimated to be 200,000 years before present (44). Haplogroup D is obviously much younger, which indicates that positive selection was at work in a period considerably postdating the emergence of anatomically modern humans in Africa. We note that the age of haplogroup D coincides with the introduction of anatomically modern humans into Europe about 40,000 years ago, as well as the dramatic shift in the archeological record indicative of modern human behavior, such as art and the use of symbolism (i.e., the "Upper Paleolithic revolution") (45).

If haplogroup D indeed experienced a recent selective sweep, it should show low poly-
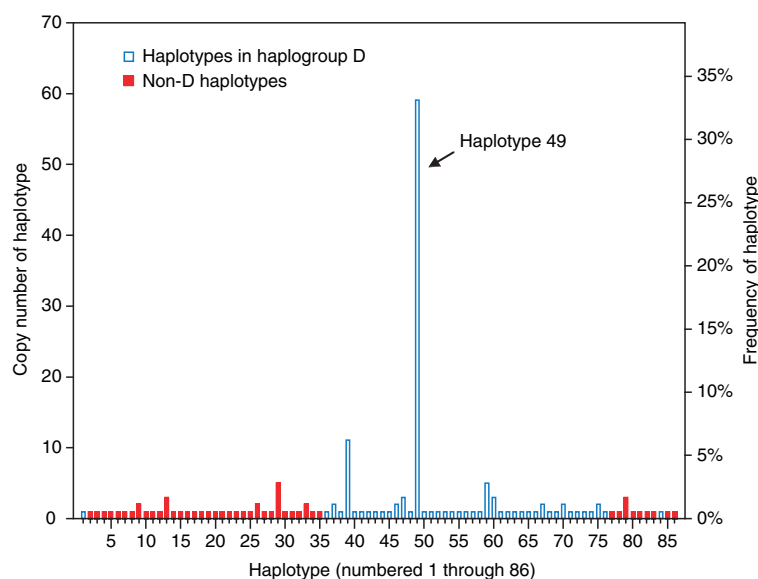

Fig. 2. Frequencies of 86 inferred *Microcephalin* haplotypes in the 89-individual Coriell panel. Haplotypes in haplogroup D are indicated by blue-edged bars; non-D haplotypes are indicated by solid red bars.

morphism and an excess of rare alleles (46). To confirm this, we calculated nucleotide diversity (π) and Tajima's D for the 47 individuals who are homozygous for haplogroup D chromosomes, and we compared these values to those of the non-D chromosomes. The π value of the D chromosomes is lower, by a factor of 12, than that of the non-D chromosomes (0.000077 and 0.00092, respectively), even though the D chromosomes represent about 70% of the chromosomes in the panel. Tajima's D, which is a summary statistic for the frequency spectrum of alleles, is –2.3 for haplogroup D (whereas it is –1.2 for the non-D chromosomes). This strongly negative Tajima's D indicates a starlike genealogy for haplogroup D chromosomes (47). Thus, both summary statistics contrast sharply between D and non-D chromosomes and are consistent with the recent age and rapid expansion of haplogroup D. We note that these calculations do not provide a statistically stringent test of positive selection, because they are done on subsets of the genealogy. Nevertheless, they do

reveal qualitative signatures of positive selection that further corroborate the more stringent statistical tests described earlier.

Another sign of a positive selective sweep is extended LD around the selected allele. This is apparent in the region of *Microcephalin* investigated here, where haplogroup D chromosomes show near-complete LD across the entire region. The only exceptions are haplotypes 1, 68, and 84 (each found in a single copy in the Coriell panel), which are recombinants between D and non-D chromosomes as evidenced by recombination tracts (table S2). The remaining 121 copies of haplogroup D chromosomes show no evidence of recombination. By comparison, the non-D chromosomes do not display any significant LD across the region.

To probe the extent of LD beyond the 29-kb core region, we sequenced the Coriell panel for two segments of about 3 kb each, situated at the beginning and end of the gene separated from each other by about 235 kb. In these flanking regions, there is clear evidence

of LD decay from the core region, which supports the idea that selection has most likely operated on a site (or sites) around the core region. Our present data cannot resolve the exact site(s) of selection, and the G37995C nonsynonymous SNP used to define haplogroup D is just a candidate.

To obtain a more detailed frequency distribution of haplogroup D across the globe, we analyzed a much larger human population panel containing 1184 globally diverse individuals. We genotyped the diagnostic G37995C SNP in this panel to infer the frequency of haplogroup D chromosomes (Fig. 3). Geographic variation was observed, with sub-Saharan populations generally having lower frequencies than others. The statistic for genetic differentiation, F_{ST}, is 0.48 between sub-Saharans and others, which indicates strong differentiation (48) and is significantly higher than the genome average of 0.12 ($P < 0.03$ based on previously established genomewide F_{ST} distribution) (49). Such population differentiation may reflect a Eurasian origin of haplogroup D, local adaptation, and/or
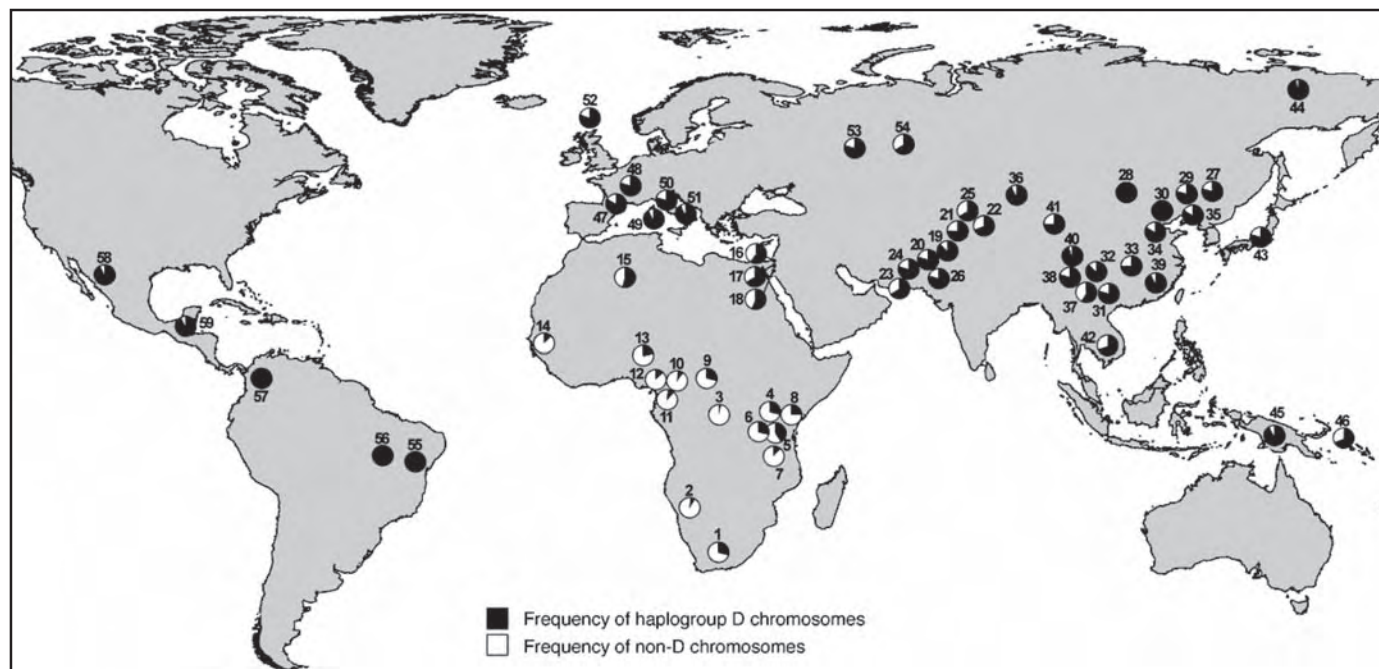

Fig. 3. Global frequencies of *Microcephalin* haplogroup D chromosomes (defined as having the derived C allele at the G37995C diagnostic SNP) in a panel of 1184 individuals. For each population, the country of origin, number of individuals sampled, and frequency of haplogroup D chromosomes are given (in parentheses) as follows: 1, Southeastern and Southwestern Bantu (South Africa, 8, 31.3%); 2, San (Namibia, 7, 7.1%); 3, Mbuti Pygmy (Democratic Republic of Congo, 15, 3.3%); 4, Masai (Tanzania, 27, 29.6%); 5, Sandawe (Tanzania, 32, 39.1%); 6, Burunge (Tanzania, 28, 30.4%); 7, Turu (Tanzania, 23, 15.2%); 8, Northeastern Bantu (Kenya, 12, 25%); 9, Biaka Pygmy (Central African Republic, 32, 26.6%); 10, Zime (Cameroon, 23, 8.7%); 11, Bakola Pygmy (Cameroon, 24, 10.4%); 12, Bamoun (Cameroon, 28, 17.9%); 13, Yoruba (Nigeria, 25, 24%); 14, Mandenka (Senegal, 24, 16.7%); 15, Mozabite [Algeria (Mzab region), 29, 53.5%]; 16, Druze [Israel (Carmel region), 44, 60.2%]; 17, Palestinian [Israel (Central), 40, 63.8%]; 18, Bedouin [Israel (Negev region), 44, 54.6%]; 19, Hazara (Pakistan, 20, 85%); 20, Balochi (Pakistan, 23, 78.3%); 21, Pathan (Pakistan, 23, 76.1%); 22, Burusho (Pakistan, 25, 66%); 23, Makrani (Pakistan, 24, 62.5%); 24, Brahui (Pakistan, 25, 78%); 25, Kalash (Pakistan, 24, 62.5%); 26, Sindhi (Pakistan, 25, 78%); 27, Hezhen (China, 9, 77.8%); 28, Mongola (China, 10, 100%); 29, Daur (China, 10, 85%); 30, Orogen (China, 10, 100%); 31, Miaozu (China, 9, 77.8%); 32, Yizu (China, 10, 85%); 33, Tujia (China, 10, 75%); 34, Han (China, 41, 82.9%); 35, Xibo (China, 9, 83.3%); 36, Uygur (China, 10, 90%); 37, Dai (China, 9, 55.6%); 38, Lahu (China, 10, 85%); 39, She (China, 9, 88.9%); 40, Naxi (China, 10, 95%); 41, Tu (China, 10, 75%); 42, Cambodian (Cambodia, 11, 72.7%); 43, Japanese (Japan, 27, 77.8%); 44, Yakut [Russia (Siberia region), 25, 98%]; 45, Papuan (New Guinea, 17, 91.2%); 46, NAN Melanesian (Bougainville, 18, 72.2%); 47, French Basque (France, 24, 83.3%); 48, French (France, 28, 78.6%); 49, Sardinian (Italy, 26, 90.4%); 50, North Italian [Italy (Bergamo region), 13, 76.9%]; 51, Tuscan (Italy, 8, 87.5%); 52, Orcadian (Orkney Islands, 16, 81.3%); 53, Russian (Russia, 24, 79.2%); 54, Adygei [Russia (Caucasus region), 15, 63.3%]; 55, Karitiana (Brazil, 21, 100%); 56, Surui (Brazil, 20, 100%); 57, Colombian (Colombia, 11, 100%); 58, Pima (Mexico, 25, 92%); 59, Maya (Mexico, 25, 92%).

demographic factors such a bottleneck associated with human migration out of Africa 50,000 to 100,000 years ago.

Previous studies have shown that *Microcephalin* is a specific regulator of brain size (*13*, *14*) and that this gene has evolved under strong positive selection in the primate lineage leading to *Homo sapiens* (*7*, *8*). Here, we present compelling evidence that *Microcephalin* has continued its trend of adaptive evolution beyond the emergence of anatomically modern humans. The specific function of *Microcephalin* in brain development makes it likely that selection has operated on the brain. Yet, it remains formally possible that an unrecognized function of *Microcephalin* outside of the brain is actually the substrate of selection. If selection indeed acted on a brain-related phenotype, there could be several possibilities, including brain size, cognition, personality, motor control, or susceptibility to neurological and/or psychiatric diseases. We hypothesize that D and non-D haplotypes have different effects on the proliferation of neural progenitor cells, which in turn leads to different phenotypic outcomes of the brain visible to selection.

References and Notes

1. J. N. Spuhler, *The Evolution of Man's Capacity for Culture* (Wayne State Univ. Press, Detroit, MI, 1959).
2. J. H. Jerison, *Evolution of the Brain and Intelligence* (Academic Press, New York, 1973).
3. W. Enard et al., *Nature* **418**, 869 (2002).
4. J. Zhang, *Genetics* **165**, 2063 (2003).
5. P. D. Evans et al., *Hum. Mol. Genet.* **13**, 489 (2004).
6. N. Kouprina et al., *PLoS Biol.* **2**, E126 (2004).
7. Y. Q. Wang, B. Su, *Hum. Mol. Genet.* **13**, 1131 (2004).
8. P. D. Evans, J. R. Anderson, E. J. Vallender, S. S. Choi, B. T. Lahn, *Hum. Mol. Genet.* **13**, 1139 (2004).
9. R. J. Ferland et al., *Nat. Genet.* **36**, 1008 (2004).
10. H. H. Stedman et al., *Nature* **428**, 415 (2004).
11. F. Burki, H. Kaessmann, *Nat. Genet.* **36**, 1061 (2004).
12. S. Dorus et al., *Cell* **119**, 1027 (2004).
13. A. P. Jackson et al., *Am. J. Hum. Genet.* **63**, 541 (1998).
14. A. P. Jackson et al., *Am. J. Hum. Genet.* **71**, 136 (2002).
15. W. B. Dobyns, *Am. J. Hum. Genet.* **112**, 315 (2002).
16. G. H. Mochida, C. A. Walsh, *Curr. Opin. Neurol.* **14**, 151 (2001).
17. C. G. Woods, J. Bond, W. Enard, *Am. J. Hum. Genet.* **76**, 717 (2005).
18. E. Roberts et al., *Eur. J. Hum. Genet.* **7**, 815 (1999).
19. L. Moynihan et al., *Am. J. Hum. Genet.* **66**, 724 (2000).
20. C. R. Jamieson, C. Govaerts, M. J. Abramowicz, *Am. J. Hum. Genet.* **65**, 1465 (1999).
21. J. Bond et al., *Nat. Genet.* **32**, 316 (2002).
22. G. F. Leal et al., *J. Med. Genet.* **40**, 540 (2003).
23. J. Bond et al., *Nat. Genet.* **37**, 353 (2005).
24. T. Huyton, P. A. Bates, X. Zhang, M. J. Sternberg, P. S. Freemont, *Mutat. Res.* **460**, 319 (2000).
25. S. Y. Lin, S. J. Elledge, *Cell* **113**, 881 (2003).
26. X. Xu, J. Lee, D. F. Stern, *J. Biol. Chem.* **279**, 34091 (2004).
27. M. Trimborn et al., *Am. J. Hum. Genet.* **75**, 261 (2004).
28. S. L. Gilbert, W. B. Dobyns, B. T. Lahn, *Nat. Rev. Genet.* **6**, 581 (2005).
29. M. Stephens, N. J. Smith, P. Donnelly, *Am. J. Hum. Genet.* **68**, 978 (2001).
30. M. Stephens, P. Donnelly, *Am. J. Hum. Genet.* **73**, 1162 (2003).
31. S. A. Tishkoff et al., *Science* **293**, 455 (2001).
32. P. C. Sabeti et al., *Nature* **419**, 832 (2002).
33. E. Wang et al., *Am. J. Hum. Genet.* **74**, 931 (2004).
34. T. Bersaglieri et al., *Am. J. Hum. Genet.* **74**, 1111 (2004).
35. E. E. Thompson et al., *Am. J. Hum. Genet.* **75**, 1059 (2004).
36. H. Stefansson et al., *Nat. Genet.* **37**, 129 (2005).
37. R. R. Hudson, *Oxf. Surv. Evol. Biol.* **7**, 1 (1990).
38. R. R. Hudson, *Bioinformatics* **18**, 337 (2002).
39. A. Kong et al., *Nat. Genet.* **31**, 241 (2002).
40. E. Zietkiewicz et al., *J. Mol. Evol.* **47**, 146 (1998).
41. H. Harpending, A. Rogers, *Annu. Rev. Genomics Hum. Genet.* **1**, 361 (2000).
42. H. Tang, D. O. Siegmund, P. Shen, P. J. Oefner, M. W. Feldman, *Genetics* **161**, 447 (2002).
43. R. Thomson, J. K. Pritchard, P. Shen, P. J. Oefner, M. W. Feldman, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7360 (2000).
44. I. McDougall, F. H. Brown, J. G. Fleagle, *Nature* **433**, 733 (2005).
45. R. G. Klein, *The Human Career: Human Biological and Cultural Origins* (Univ. of Chicago Press, Chicago, 1999).
46. M. Bamshad, S. P. Wooding, *Nat. Rev. Genet.* **4**, 99 (2003).
47. F. Tajima, *Genetics* **123**, 585 (1989).
48. S. Wright, *Evolution and the Genetics of Populations* (Univ. of Chicago Press, Chicago, 1978).
49. J. M. Akey, G. Zhang, K. Zhang, L. Jin, M. D. Shriver, *Genome Res.* **12**, 1805 (2002).
50. We thank the Coriell Institute for Medical Research, the Centre d'Etude du Polymorphisme Humain (CEPH), and A. Froment for human DNA samples. We thank H. M. Cann, S. Dorus, E. E. Eichler, N. M. Pearson, A. Di Rienzo, M. Kreitman, and J. K. Pritchard for technical support and/or helpful discussions. Supported in part by the Searle Scholarship and the Burroughs Wellcome Career Award (to B.T.L.), and David and Lucile Packard Career Award, the Burroughs Wellcome Career Award, and NSF grant BCS-0196183 (to S.A.T).

Ongoing Adaptive Evolution of *ASPM*, a Brain Size Determinant in *Homo sapiens*

Nitzan Mekel-Bobrov,[1,2] Sandra L. Gilbert,[1] Patrick D. Evans,[1,2] Eric J. Vallender,[1,2] Jeffrey R. Anderson,[1] Richard R. Hudson,[3] Sarah A. Tishkoff,[4] Bruce T. Lahn[1]*

The gene *ASPM* (*abnormal spindle-like microcephaly associated*) is a specific regulator of brain size, and its evolution in the lineage leading to *Homo sapiens* was driven by strong positive selection. Here, we show that one genetic variant of *ASPM* in humans arose merely about 5800 years ago and has since swept to high frequency under strong positive selection. These findings, especially the remarkably young age of the positively selected variant, suggest that the human brain is still undergoing rapid adaptive evolution.

Homozygous null mutations of *ASPM* cause primary microcephaly, a condition characterized by severely reduced brain size with otherwise normal neuroarchitecture (*1*). Studies

[1]Howard Hughes Medical Institute, Department of Human Genetics, [2]Committee on Genetics, [3]Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637, USA. [4]Department of Biology, University of Maryland, College Park, MD 20742, USA.

*To whom correspondence should be addressed. E-mail: blahn@bsd.uchicago.edu

have suggested that *ASPM* may regulate neural stem cell proliferation and/or differentiation during brain development, possibly by mediating spindle assembly during cell division (*1*, *2*). Phylogenetic analysis of *ASPM* has revealed strong positive selection in the primate lineage leading to *Homo sapiens* (*3–5*), especially in the past 6 million years of hominid evolution in which ASPM acquired about one advantageous amino acid change every 350,000 years (*4*). These data argue that *ASPM*

may have contributed to human brain evolution (*3–6*). Here, we investigate whether positive selection has continued to operate on *ASPM* since the emergence of anatomically modern humans.

Human *ASPM* has 28 exons with a 10,434–base pair open reading frame (*1*) (fig. S1). We resequenced the entire 62.1-kb genomic region of *ASPM* in samples from 90 ethnically diverse individuals obtained through the Coriell Institute and from a common chimpanzee (*7*). This revealed 166 polymorphic sites (table S1). Using established methodology (*7*), we identified 106 haplotypes. One haplotype, numbered 63, had an unusually high frequency of 21%, whereas the other haplotypes ranged from 0.56% to 3.3% (fig. S2). Moreover, this haplotype differed consistently from the others at multiple polymorphic sites (save for a few rare haplotypes that are minor mutational or recombinational variants of haplotype 63, as discussed later) (table S2). Two of these polymorphic sites are nonsynonymous, both in exon 18, and are denoted A44871G and C45126A (numbers indicate genomic positions from the start codon, and letters at the beginning and end indicate ancestral and derived alleles, respectively). These two sites reside in a region of the open reading frame that was shown previously to have experienced par-
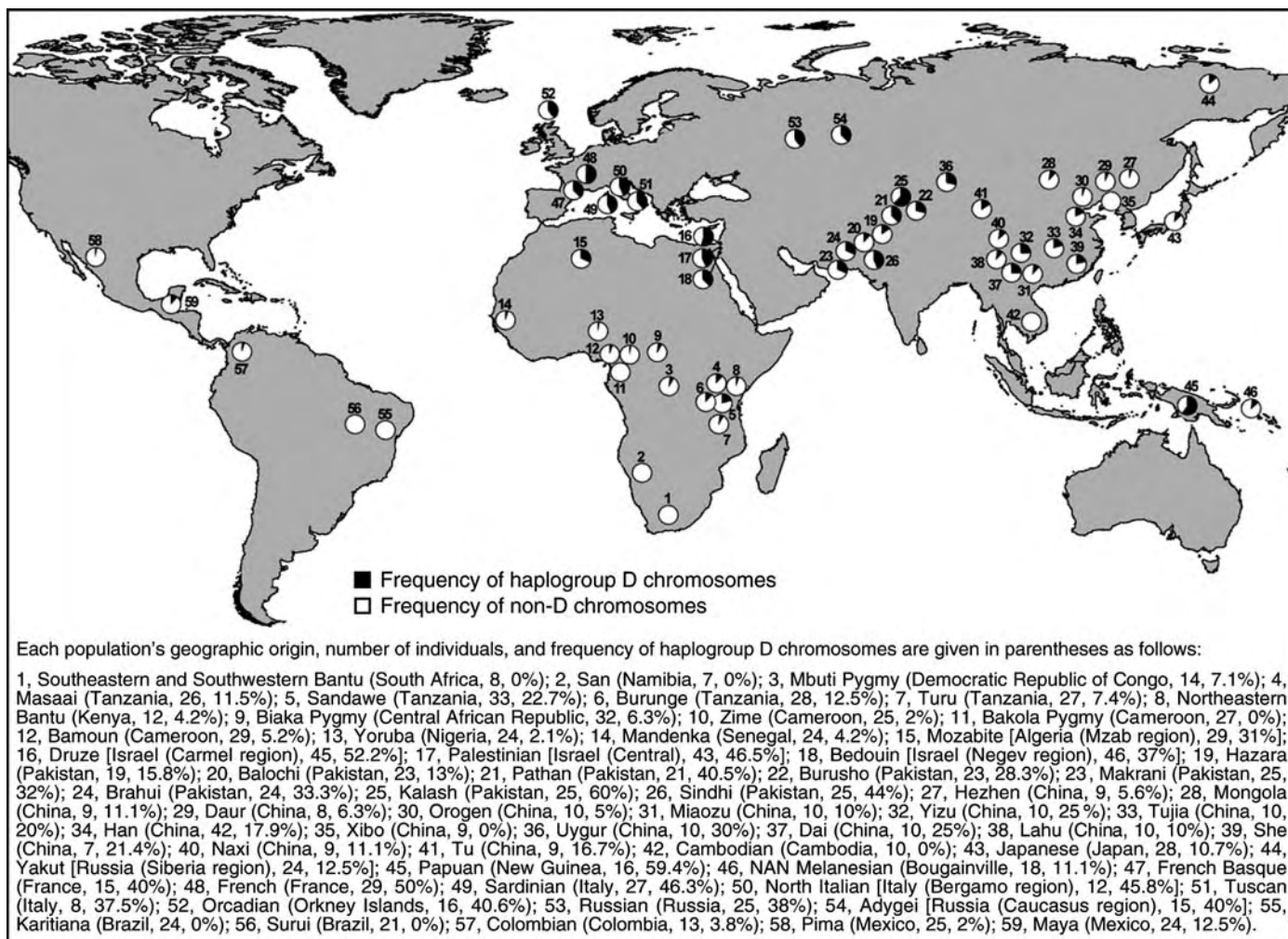
Fig. 1. Worldwide frequencies of *ASPM* haplogroup D chromosomes (defined as having the derived G allele at the A44871G diagnostic polymorphism), based on a panel of 1186 individuals.

ticularly strong positive selection in the lineage leading to humans (*4*) (fig. S1).

The unusually high frequency of haplotype 63 is strongly suggestive of positive selection (*8*). We tested the statistical significance of the possibility of positive selection using coalescent modeling (*7*). The frequency of haplotype 63 is notably higher in Europeans and Middle Easterners (including Iberians, Basques, Russians, North Africans, Middle Easterners, and South Asians), as compared with other populations (table S1). We therefore focused on this group to take advantage of its relatively simple and homogeneous demographic structure (*9*). Because 7 of the 50 Europeans and Middle Easterners were homozygous for haplotype 63, we tested the probability of obtaining 7 or more homozygotes (among 50) for a single haplotype across a 62.1-kb region containing 122 segregating sites (the number of polymorphic sites found in Europeans and Middle Easterners). The recombination rate and the gene conversion rate of the locus used in the test were obtained from our polymorphism data (*7*). For demographic history, we

assumed a severe bottleneck followed by exponential growth (*7*) that is likely to be much more stringent than the bottleneck associated with the colonization of Eurasia (*10*). These parameters produced a highly significant departure from the neutral expectation ($P <$ 0.00001). The simulation was repeated with a wide range of demographic histories (*7*), all of which produced highly significant results. We repeated the above tests on the entire Coriell panel, which also demonstrated strong significance. Finally, we repeated these tests using the inferred frequency of haplotype 63 (instead of the number of individuals who are homozygous for this haplotype) and again obtained significant statistical values. These data indicate that haplotype 63 has spread to high frequency under positive selection. The two nonsynonymous polymorphisms previously mentioned are either the target of selection or closely linked to the target site.

We define haplogroup D (where D stands for "derived") as the class of haplotypes with the derived G allele at the A44871G nonsynonymous polymorphic site. The haplotypes

with the A allele are defined as non-D. This classification is meant to capture the notable structure in the 106 haplotypes. Haplogroup D comprises two subgroups. One contains the predominant haplotype 63 and its closely related mutational variants [including haplotypes 64, 66, and 71 (table S2)]. The other contains recombinant haplotypes between haplotype 63 (or its close mutational variants) and non-D haplotypes (including haplotypes 62, 65, and 105). The frequency of haplogroup D chromosomes is 28% in the entire Coriell panel and 44% in Europeans and Middle Easterners. Another prominent feature separating D and non-D haplotypes is the fact that the two classes have fixed differences relative to each other at multiple sites, where D haplotypes have the derived alleles (except for the rare recombinants between D and non-D chromosomes). This unusual haplotype structure is consistent with the following evolutionary history: (i) a rapid increase of haplotype 63 from a single ancestral copy to high frequency, and (ii) the introduction of minor variants of haplotype 63 by mutation and by recombination between D

and non-D chromosomes (these variants make up the other members of haplogroup D).

Another signature of positive selection is extended linkage disequilibrium (LD) (8), which is evident in the 62.1-kb region. Indeed, the 50 haplogroup D chromosomes show nearly complete LD across the region, with only three cases of LD breakdown [haplotypes 62, 65, and 105, present in one, two, and one copy, respectively (table S2)]. In contrast, the non-D chromosomes do not show any unusual LD across the region. We investigated the decay of LD beyond the 62.1-kb region by sequencing the Coriell panel for two flanking intergenic segments of roughly 5 kb each, positioned 25 kb upstream and downstream of *ASPM*. There is notable LD breakdown in these two segments, confirming that the site of selection is most likely within *ASPM*.

Following established methodology (7), we estimated the coalescence age (i.e., time to the most recent common ancestor) of haplogroup D at 5800 years, with a 95% confidence interval between 500 and 14,100 years. In comparison, the coalescence age of all the chromosomes (both D and non-D) is ~800,000 years. Thus, the age of haplogroup D substantially postdates the emergence of anatomically modern humans, estimated at ~200,000 years ago (11). A rough calculation showed that the fitness advantage of haplogroup D is in the range of a few percent over non-D haplotypes.

By genotyping the A44871G diagnostic polymorphism described earlier, we obtained the global frequency distribution of haplogroup D chromosomes from a separate panel of 1186 individuals (7). Consistent with the Coriell panel, we observed much higher frequency of haplogroup D chromosomes in Europeans and Middle Easterners than in other populations (Fig. 1). The corresponding estimate of F_{ST}, a statistic of genetic differentiation, is 0.29 between Europeans/Middle Easterners and other populations and 0.31 between Europeans/Middle Easterners and sub-Saharan Africans. These values indicate considerable genetic differentiation at this locus (12). Several scenarios may account for such notable differentiation. One is that haplogroup D first arose somewhere in Eurasia and is still in the process of spreading to other regions. The other is that it arose in sub-Saharan Africa, but reached higher frequency outside of Africa partly because of the bottleneck during human migration out of Africa. Finally, it is possible that differential selective pressure in different geographic regions is partly responsible.

Collectively, our data offer strong evidence that haplogroup D emerged very recently and subsequently rose to high frequency under strong positive selection. The recent selective history of *ASPM* in humans thus continues the trend of positive selection that has operated at this locus for millions of years in the hominid lineage (3–5). Although the age of haplogroup D and its geographic distribution across Eurasia roughly coincide with two important events in the cultural evolution of Eurasia—namely, the emergence and spread of domestication from the Middle East ~10,000 years ago (13) and the rapid increase in population associated with the development of cities and written language 5000 to 6000 years ago around the Middle East (14)—the significance of this correlation is not yet clear.

References and Notes

1. J. Bond *et al.*, *Nat. Genet.* **32**, 316 (2002).
2. P. Ripoll, S. Pimpinelli, M. M. Valdivia, J. Avila, *Cell* **41**, 907 (1985).
3. J. Zhang, *Genetics* **165**, 2063 (2003).
4. P. D. Evans *et al.*, *Hum. Mol. Genet.* **13**, 489 (2004).
5. N. Kouprina *et al.*, *PLoS Biol.* **2**, E126 (2004).
6. S. L. Gilbert, W. B. Dobyns, B. T. Lahn, *Nat. Rev. Genet.* **6**, 581 (2005).
7. Materials and methods are available as supporting material on *Science* Online.
8. M. Bamshad, S. P. Wooding, *Nat. Rev. Genet.* **4**, 99 (2003).
9. N. A. Rosenberg *et al.*, *Science* **298**, 2381 (2002).
10. E. Zietkiewicz *et al.*, *J. Mol. Evol.* **47**, 146 (1998).
11. I. McDougall, F. H. Brown, J. G. Fleagle, *Nature* **433**, 733 (2005).
12. S. Wright, *Evolution and the Genetics of Populations* (Univ. of Chicago Press, Chicago, 1978).
13. L. L. Cavalli-Sforza, P. Menozzi, A. Piazza, *The History and Geography of Human Genes* (Princeton Univ. Press, Princeton, 1994).
14. S. D. Houston, *The First Writing: Script Invention as History and Process* (Cambridge Univ. Press, New York, 2004).
15. We thank the Coriell Institute for Medical Research, the Centre d'Etude du Polymorphisme Humain, and A. Froment for human DNA samples; and H. M. Cann, A. Di Rienzo, S. Dorus, E. E. Eichler, M. Kreitman, N. M. Pearson, and J. K. Pritchard for technical support and helpful discussions. Supported in part by the Searle Scholarship and Burroughs Wellcome Career Award (to B.T.L.) and by the David and Lucile Packard Career Award, Burroughs Wellcome Career Award, and NSF grant no. BCS-0196183 (to S.A.T).

Neutral Ecological Theory Reveals Isolation and Rapid Speciation in a Biodiversity Hot Spot

Andrew M. Latimer,[1]* John A. Silander Jr.,[1] Richard M. Cowling[2]

South Africa's Mediterranean-climate fynbos shrubland is a hot spot of species diversity, but its diversity patterns contrast strongly with other high-diversity areas, including the Amazon rain forest. With its extremely high levels of endemism and species turnover, fynbos is made up of dissimilar local communities that are species-rich but relatively poor in rare species. Using neutral ecological theory, we show that the relative species-abundance distributions in fynbos can be explained by migration rates that are two orders of magnitude lower than they are in tropical rain forests. Speciation rates, which are indexed by the "biodiversity parameter" Θ, are estimated to be higher than they are in any previously examined plant system.

Regions with many endemic resident species are typically grouped together as biodiversity "hot spots," but this label obscures fundamental differences among high-diversity systems in species-diversity patterns and in the processes that generate them. The fynbos shrubland of South Africa's Cape Floristic Region (CFR) presents a particularly strong contrast with other diverse systems such as the Amazon rain forest. Amazon rain forest tree communities are dominated by species with large ranges, most of which are locally rare (1, 2), whereas most fynbos shrub species are narrow endemics, and many can be locally abundant (1, 3). Neutral ecological theory (4) enables the quantification of diversity in these contrasting systems in terms of the mechanisms that generate and partition it: rates of speciation and rates of migration.

The CFR contains over 8,000 plant species, of which a majority (69%) are endemic to the region (2). Mediterranean-climate regions are among the most diverse systems outside the tropics, and among these, the CFR has the most species in the smallest area (5). Molecular phylogenetic studies have estimated that speciation rates in some lineages in the

[1]Department of Ecology and Evolutionary Biology, University of Connecticut, 75 North Eagleville Road, Storrs, CT 06269, USA. [2]Department of Botany, Terrestrial Ecology Research Unit, Post Office Box 77000, Nelson Mandela Metropolitan University, Port Elizabeth 6031, South Africa.

*To whom correspondence should be addressed. E-mail: latimer@uconn.edu

CFR are exceptionally high (*1, 3*). Moreover, CFR plants are thought to have very low migration rates (*6*). High levels of endemism among CFR species are associated with low observed seed-dispersal distances and short-distance dispersal modes (ant and passive dispersal) (*7*). Even among Mediterranean-climate regions, the CFR has very high rates of species turnover along habitat and geographic gradients, i.e., high β and γ diversities, respectively (*8*), which is a spatial pattern that suggests low migration rates (*5*). Amazon rain forest, in contrast, occupies a much larger area, and can contain as many as 300 tree species in a sample of 600 individuals (extremely high local, or α, diversity) (*9*). Relatively few Amazon tree species appear to be locally endemic, because even rare species have generally broad distributions, producing comparatively lower β and γ diversities (*10, 11*). Studies of tropical rain forest trees have shown them to be generally well dispersed, with long mean-dispersal distances and frequent long-distance dispersal events (*12, 13*).

Hubbell's neutral theory of biodiversity and biogeography ("neutral theory") provides mechanistic insight into the processes driving these two types of high-diversity systems (*4*). Neutral theory extends previous ecological work on quantifying and comparing diversity patterns by grounding its descriptions in interpretable ecological processes; its two core parameters represent the processes of speciation and migration (*14*). The "biodiversity parameter," Θ, represents the process of speciation within a "metacommunity," which is a very large collection of similar organisms found across a biogeographic region (*4*). This parameter is twice the product of the size of the metacommunity J_m (a very large number) and the speciation rate υ, which is the number of new species arising per birth (a very small number). Thus, Θ is proportional to the expected number of new species per generation across the metacommunity and can be used to quantify the diversity-generating capacity of a system. In terrestrial plant systems, Θ has been shown to range from ~1 in boreal forest to ~200 in tropical rain forests (*4*). Within the metacommunity are "local communities," which are groups of organisms at a particular locality that are linked to the metacommunity by migration. The migration parameter *m* is the proportion of individuals per generation in each local community that come from elsewhere in the metacommunity (*4*). Given values for *m*, Θ, and the size of the local community (*J*), neutral theory produces an expected relative species abundance distribution, called the zero-sum multinomial (ZSM) (*15, 16*), which can be fitted to sample inventory data.

Neutral dynamics can generate two extreme types of species-rich systems, depending on the amount of migration. A high-migration system

(*m* > 0.1) has species-rich local communities with high evenness in relative abundances. A low-migration system (*m* < 0.01) is a spatially structured collection of dissimilar local communities that are less species-rich because of local extinction of rare species through stochastic drift and because of correspondingly higher dominance by a few species in each local community. For the same value of Θ, a high-migration system will have higher local diversity, whereas a low-migration system will have proportionally higher species turnover among localities.

An advantage of neutral theory over alternative descriptors of biodiversity patterns is that its parameters represent easily interpreted ecological processes. However, little work has been done to produce parameter estimates for a range of systems and to assess what insight they provide about ecological dynamics. Because most previous applications of neutral theory to plant communities have used tropical rain forest data sets (*12, 13, 15, 17*), the application of the theory to real-world patterns has been confined to only a few geographic regions and a small portion of parameter space. Here we contrast the Amazon rain forest system with the CFR, which may be the clearest example of a species-rich low-migration system.

The CFR data come from three sites spanning the region—the Cederberg mountains in

the northwest, Cape Hangklip in the southwest, and the Zuurberg mountains at the eastern edge. Each of the sites in our data set is an identifiable subregion of the CFR, which consists of a range of hills or mountains and adjacent lowlands (*18*). For each site, we obtained existing floristic survey data, including hundreds of 5-m by 10-m plots, in which all vascular plants had been identified to species (Table 1) (*19–22*). From these data, we obtained the abundances of all woody plant species for each site (subshrubs, shrubs, and trees) (*23*). For the comparison to tropical rain forests, we used tree-count data from networks of large plots in the Manu National Park in Peru and Yasuní National Park in Ecuador (*10, 11*). Estimates for Barro Colorado Island (BCI) in Panama, using tree count data from the 50-ha plot there, were also included as a well-studied baseline (*13*).

Observed patterns of diversity and abundance in the CFR are very different from those documented for tropical rain forests (*2, 24*). Figure 1, which displays species relative abundances at a CFR site (the Cederberg) and an Amazon forest site (Manu National Park), shows that the CFR site has higher dominance by common species, with fewer rare species. Each CFR site contains fewer species than do typical tropical rain forest localities containing equivalent numbers of individuals, but local

Fig. 1. Rank-abundance plots for the Cederberg (CFR) and Manu (Amazon) sites. The triangles (Cederberg, red lines) and circles (Manu, blue lines) show abundance of the ranked species for each site, normalized by total abundance and log transformed. Solid lines show rank abundances predicted by neutral theory for each site, using maximum likelihood parameter estimates. Dashed lines indicate bootstrapped 95% confidence intervals around the predicted rank abundances, quantifying sampling error expected when drawing from a neutral community with known parameters.
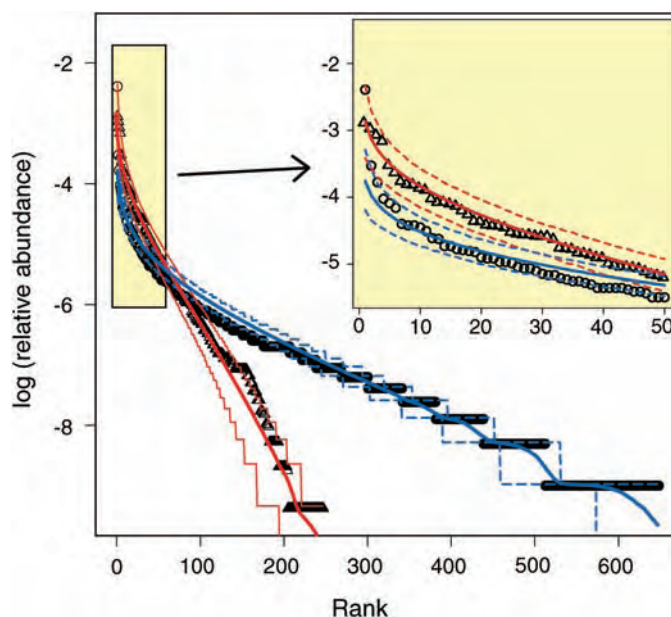
Table 1. Data sets.

Site	Plots	Species	Individuals
Cederberg (*20*)	191	247	11,765
Cape Hangklip (*19*)	251	247	24,003
Zuurberg (*21*)	69	114	8,916
Manu National Park, Peru (*10, 11*)	15	648	8,055
Yasuní National Park, Ecuador (*10, 11*)	14	546	7,613
BCI, Panama (*13*)	1	225	21,457

endemism in the CFR is extremely high, resulting in very high species turnover among sites (*18*, *25*). The mean adjusted Jaccard's Index [a measure of taxonomic similarity between sites that adjusts for sampling effects (*26*)] for the three CFR sites studied here is 0.024, compared with 0.47 for the two Amazon sites, despite the shorter distances among the CFR sites (~150 to 400 km versus ~1500 km).

We used Bayesian and maximum-likelihood methods to fit the ZSM to species-abundance data from each site (*23*). Parameter estimates were obtained by using a likelihood derived from the closed-form expression for the ZSM given in Volkov *et al.* (*15*), which provides the predicted distribution of abundances for a sample of species from a neutral community. To obtain Bayesian parameter estimates, uninformative priors were placed on the parameters (*23*).

Bayesian estimates of *m* for the CFR ranged from 0.0018 to 0.045, which is about two orders of magnitude lower than the estimates for the rain forest sites (0.54 for Yasuní, 0.55 for Manu, and 0.1 for BCI) (Fig. 2 and Table 2). The maximum likelihood estimates for the parameters were consistent with the Bayesian results (Table 2). None of the credible intervals of the CFR estimates of *m* overlapped those of the Amazon estimates. Within the CFR, the lowest migration estimate is for Cape Hangklip, which is one of the most diverse areas of the CFR (*2*, *25*). Zuurberg, at the eastern edge of the CFR, had much higher estimates of *m*, approaching those of BCI in Panama.

Estimates of Θ for the CFR were very high, peaking in the southwest (Cape Hangklip) at $\Theta = 530$, which is more than twice the highest value yet estimated for any rain forest (~200) (*4*). Because the posterior distribution for Θ in low-migration communities is right-skewed (Fig. 2D), we present the posterior mode rather than the posterior mean as a conservative estimate (Table 2). The credible interval for Θ for Cape Hangklip does not overlap the interval for Manu, the richer of the two Amazon communities, and this value is the highest estimate of Θ for any plant system yet examined. As with *m*, there is a longitudinal trend in Θ within the CFR: the highest Θ values coincide with the lowest migration rates in the southwest.

The eastern part of the CFR, including the Zuurberg site, has more-even seasonal rainfall patterns than does the southwestern CFR, and the vegetation contains a higher proportion of widely distributed, well-dispersed forest and thicket species (*27*). The lowest migration rates are thus associated with the highly seasonal winter rainfall conditions in the southwest, as well as with the highest levels of endemism and species turnover (*28*). The southwestern site (Cape Hangklip) is also geographically central, whereas the eastern site (Zuurberg) is the most peripheral, suggesting that edge effects may be reducing species diversity at the periphery of the region.

Combining the data from the subregions within each region (CFR and Amazon) produces a proxy for samples from the two metacommunities. By using the composite data, we can assess the importance of migration limitation by comparing the fits of models with and without a migration parameter (*16*). Estimates of Θ for the metacommunity neutral model with no migration limitation were performed by maximizing the likelihood of the metacommunity sampling distribution (*16*). The assumption of unlimited migration drastically lowers the estimate of Θ for the CFR (from 697 to 88.9), whereas Θ for the Amazon is only moderately reduced (251 to 219). The reduction in the CFR estimate is not surprising, because if migration limitation is ignored, the lower species richness can be accommodated only by lower values of Θ. The assumption of unlimited migration is clearly
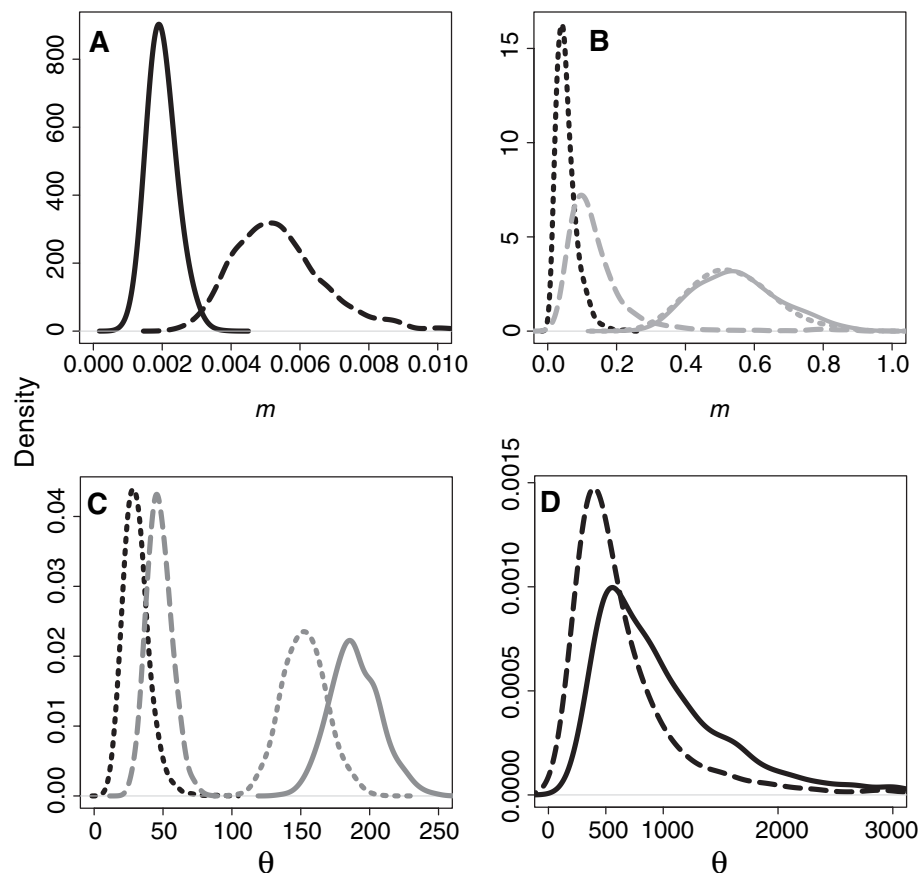

Fig. 2. Bayesian estimates of the ZSM parameters *m* (upper two panels) and Θ (lower two panels). Posterior densities for tropical rain forest sites are shown in gray and CFR sites are shown in black. (**A**) The low-*m* sites: Cape Hangklip (solid black line) and Cederberg (dashed black). (**B**) The higher-*m* sites: Zuurberg (dotted black), Barro Colorado Island (dashed gray), Yasuní (dotted gray), and Manu (solid gray). (**C**) Zuurberg (dotted black line), BCI (dashed gray line), Yasuní (dotted gray), and Manu (solid gray). (**D**) Cederberg (dashed black) and Cape Hangklip (solid black).

Table 2. Parameter estimates under neutral theory ZSM distribution. Estimates for neutral theory parameters Θ and *m* are shown. Both maximum likelihood estimates (MLE) and Bayesian estimates [posterior (Post.) mode] are presented; the 0.025 and 0.975 quantiles for the Bayesian estimates quantify the uncertainty associated with the estimates.

Site	Θ				*m*			
	MLE	Post. mode	0.025 quantile	0.975 quantile	MLE	Post. Mode	0.025 quantile	0.975 quantile
Cederberg	394.6	370	198.9	2350.5	0.0055	0.0053	0.0032	0.0090
Hangklip	640.3	530	328.3	3033.7	0.0020	0.0018	0.0014	0.0029
Zuurberg	30.2	29	15.3	52.5	0.037	0.045	0.015	0.13
Manu	187.4	183	154.9	228.3	0.53	0.55	0.34	0.81
Yasuní	152.8	153	121.9	188.6	0.50	0.50	0.34	0.80
BCI	47.7	45	31.3	67.0	0.093	0.10	0.048	0.44

unsatisfactory for both systems, because a likelihood ratio test strongly favors the model with migration limitation in both cases [for Amazon: $\chi^2 = 24.6$, 1 degree of freedom (d.f.), $P < 0.001$; for CFR: $\chi^2 = 19.4$, 1 d.f., $P < 0.001$].

Estimates of Θ can be converted to estimates of per capita speciation rate if it is possible to estimate the number of individuals in the metacommunity. With a much larger area, the Amazon rain forest metacommunity almost certainly contains at least as many individuals as does fynbos, which covers <50,000 km². The high estimates for Θ thus imply a substantially higher per capita speciation rate in the CFR. These results are consistent with the prevailing view that the western CFR is an extremely migration-limited system with extraordinarily high speciation rates. Each of the local communities in our data set is an identifiable subregion of the CFR that consists of a range of hills or mountains and adjacent lowlands. Thus, the CFR metacommunity is topographically fragmented, which predicts low migration rates, consistent with our results. With these low migration rates, it is likely that local communities will be sufficiently isolated to allow ecological drift to cause divergence among communities, because few individuals per generation will be exchanged (19). Further, populations of individual species will be genetically isolated, so that genetic drift will tend to cause divergence and the formation of

new species. Thus, our results support the view that the fynbos metacommunity consists of topographical islands isolated not by water but by drier lowlands. This pattern contrasts sharply with that found in tropical rain forests, which exhibit high connectivity over long distances. Within the CFR, the association of the highest values of Θ, and thus speciation rates, with the lowest migration rates supports the view that isolation over short spatial scales (1 to 100 km) has played a role in generating, as well as structuring, the high diversity of the CFR.

References and Notes

1. H. P. Linder, C. R. Hardy, *Philos. Trans. R. Soc. London Ser. B* **359**, 1623 (2004).
2. P. Goldblatt, J. Manning, *Cape Plants: A Conspectus of the Cape Flora of South Africa* (National Botanical Institute of South Africa, Cape Town, 2000).
3. J. E. Richardson et al., *Nature* **412**, 181 (2001).
4. S. P. Hubbell, *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, Princeton, 2001).
5. R. M. Cowling, P. W. Rundel, B. B. Lamont, M. K. Arroyo, M. Arianoutsou, *Trends Ecol. Evol.* **11**, 362 (1996).
6. P. Slingsby, W. J. Bond, *S. Afr. J. Bot.* **51**, 30 (1985).
7. D. J. McDonald, J. M. Juritz, R. M. Cowling, W. J. Knottenbelt, *Plant Syst. Evol.* **195**, 137 (1995).
8. R. H. Whittaker, *Taxon* **21**, 213 (1972).
9. H. ter Steege et al., *Biodiversity Conserv.* **12**, 2255 (2003).
10. N. C. A. Pitman, J. W. Terborgh, M. R. Silman, P. V. Nuñez, *Ecology* **80**, 2651 (1999).
11. N. C. A. Pitman et al., *Ecology* **82**, 2101 (2001).
12. O. J. Hardy, B. Sonké, *For. Ecol. Manage.* **197**, 191 (2004).
13. R. Condit et al., *Science* **295**, 666 (2002).
14. J. Chave, *Ecol. Lett.* **7**, 241 (2004).
15. I. Volkov, J. R. Banavar, S. P. Hubbell, A. Maritan, *Nature* **424**, 1035 (2003).
16. D. Alonso, A. J. McKane, *Ecol. Lett.* **7**, 901 (2004).
17. B. J. McGill, *Nature* **422**, 881 (2003).
18. R. M. Cowling, Ed., *The Ecology of Fynbos: Nutrients, Fire and Diversity* (Oxford Univ. Press, Cape Town, 1992).
19. C. Boucher, *Bothalia* **12**, 455 (1978).
20. H. C. Taylor, *Cederberg Vegetation and Flora* (National Botanical Institute, Cape Town, 1996).
21. B. E. Van Wyk, P. A. Novellie, C. M. Van Wyk, *Bothalia* **18**, 211 (1988).
22. M. J. A. Werger, *Bothalia* **11**, 309 (1974).
23. Materials and methods are available as supporting material on *Science* Online.
24. R. M. Cowling, B. B. Lamont, *Aust. J. Bot.* **46**, 335 (1998).
25. R. M. Cowling, P. M. Holmes, *Biol. J. Linn. Soc.* **47**, 367 (1992).
26. A. Chao, R. L. Chazdon, R. K. Colwell, T.-J. Shen, *Ecol. Lett.* **8**, 148 (2005).
27. R. M. Cowling, S. Proches, in *Plant Diversity and Complexity Patterns. Local, Regional and Global Dimensions*, H. Balslev, Ed. (The Royal Danish Academy of Sciences and Letters, Copenhagen, 2005), vol. 55, pp. 273–288.
28. R. M. Cowling, A. T. Lombard, *Diversity Distr.* **8**, 163 (2002).
29. Supported by NSF grant DEB008901 to J.A.S. and R.M.C. and by an NSF Graduate Research Fellowship to A.M.L. We are grateful to L. Mucina and S. Proches for assistance with CFR data sets, N. Pitman for providing Amazon data, P. Holmes Rebelo for vegetation survey data, and A. Rebelo and the Protea Atlas Project of the South African National Biodiversity Institute for interpretation. I. Koltracht provided code for numerical integration and B. McGill, R. Colwell, P. Linder, and three anonymous reviewers provided helpful comments.

Generating Electricity While Walking with Loads

Lawrence C. Rome,[1,2]* Louis Flynn,[1] Evan M. Goldman,[1] Taeseung D. Yoo[1]

We have developed the suspended-load backpack, which converts mechanical energy from the vertical movement of carried loads (weighing 20 to 38 kilograms) to electricity during normal walking [generating up to 7.4 watts, or a 300-fold increase over previous shoe devices (20 milliwatts)]. Unexpectedly, little extra metabolic energy (as compared to that expended carrying a rigid backpack) is required during electricity generation. This is probably due to a compensatory change in gait or loading regime, which reduces the metabolic power required for walking. This electricity generation can help give field scientists, explorers, and disaster-relief workers freedom from the heavy weight of replacement batteries and thereby extend their ability to operate in remote areas.

Over the past century, humans have become increasingly dependent on technology, particularly electronic devices. During the past decade, electronic devices have become more mobile, enabling people to use medical, communication, and Global Positioning System (GPS) devices as they move around cities or in the wilderness. At present, all of these devices are powered by batteries, which have a limited energy storage capacity and add considerable weight. Although substantial progress has been made in reducing the power requirements of devices and increasing the power densities of batteries, there has not been a breakthrough in the parallel development of a portable and renewable human-driven energy source (1, 2).

The combination of limited energy and the large weight of batteries poses the most critical problem for individuals having high electricity demands in remote areas and who are already carrying heavy loads (such as field scientists or explorers on prolonged expeditions). At present, replacement batteries may make up a substantial proportion (as much as 25%) of the very heavy packs (>36 kg or 80 lbs) that such users must carry (3). To help solve this problem, we developed a passive device, the suspended-load backpack, which extracts mechanical energy from the vertical movement of the load during walking and converts it to electricity for powering portable devices.

During terrestrial locomotion, the environment does no work on the body (except for the small force of aerodynamic drag) and conversely, humans do no work on the environment. Rather, almost all of the mechanical work is generated and dissipated inside the body (4, 5). This makes it exceedingly difficult to capture mechanical energy to drive an electrical energy conversion apparatus, because the device would need to be either surgically placed within the body or attached to the outside of the body (such as an exoskeleton), which would affect the person's maneuverability and comfort. Therefore, researchers in the field have focused on putting devices in the only acces-

[1]Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA. [2]Marine Biological Laboratory, Woods Hole, MA 02543, USA.

*To whom correspondence should be addressed. E-mail: lrome@sas.upenn.edu

sible location: the shoe. Such "heel-strike" devices, however, have permitted only small levels of electrical energy generation (10 to 20 mW) (2, 6). The primary reason for this limitation is that on a hard surface, essentially no mechanical work (force times distance) is done at the foot/ground contact point, because under normal circumstances the point of vertical force application does not move in the vertical plane (that is, distance = ~0).

Although one can make the shoe compliant so that the foot moves a small distance because of compression of the sole and heel (7), this is problematic because increasing compliance leads to declining maneuverability and stability. Although considerable effort has gone into developing exotic energy-generating technologies for shoe devices (8), the small magnitude of the mechanical energy source remains a limitation.

We recognized that the vertical movement of a heavy load in the gravitational field during walking represents a heretofore untapped source of mechanical energy and a potential opportunity to generate substantial levels of electricity. During walking, a person moves like an inverted pendulum (4, 5, 9): One foot is put down and then the body vaults over it, causing the hip to move up and down by 4 to 7 cm (10) (Fig. 1). Thus, if one is carrying a load in a backpack, because it is fixed to the body, it has to go up and down the same vertical distance (Fig. 1). A considerable amount of mechanical energy must be transferred (or generated de novo by the muscles) if the load is heavy. In the case of a 36-kg load, 18 J of mechanical energy transfer (or work) accompanies each step (assuming 5 cm displacement), and at two steps s^{-1}, this is equivalent to 35 W. Although this represents a large potential source of mechanical energy, it is also inaccessible if the load is rigidly attached to the body. We reasoned that decoupling the load from the body would allow the differential movement (between the load and the body) necessary for mechanical energy extraction and ultimately electricity production. We therefore designed a device, the suspended-load backpack (Fig. 2), that could be interposed between the body and the load, resulting in differential movement (Fig. 2) and the potential for generating a considerable amount of electrical energy.

Figure 3 shows the displacement and electrical output from the generator of a person walking with a 38-kg load (11). In this trial, the relative movement of the load with respect to the pack frame was approximately 4.5 cm (top panel). The linear velocity of the rack, in turn, drove the generator (a 25:1 geared dc motor) up to ~5000 rpm. The middle panel shows the voltage output of the generator. In these experiments, the output of the generator ran through a fixed "load" resistor (25 ohms), and hence the electrical power, calculated as voltage^2/resistance, is shown in the bottom panel. The

average electrical power in this trace is 5.6 W. This determination of electrical power was confirmed by joule heating experiments (11).

Six male participants walked at speeds ranging from of 4.0 to 6.4 km hour^{-1} (2.5 to 4.0 mph) while carrying 20-, 29-, and 38-kg loads in addition to the fixed portion of the instrumented pack frame, which weighed 5.6 kg (12). Average electrical power increased

with walking speed and generally increased with the weight of the load in the pack (Fig. 4). Further, while walking up a 10% incline, electrical power generation for a given load and speed was equal to or greater than that on the flat (13). The maximum electrical power output obtained on the flat was 7.37 W (±SE = 0.49, n = 6 participants), or about 300 times higher than previously published values gen-
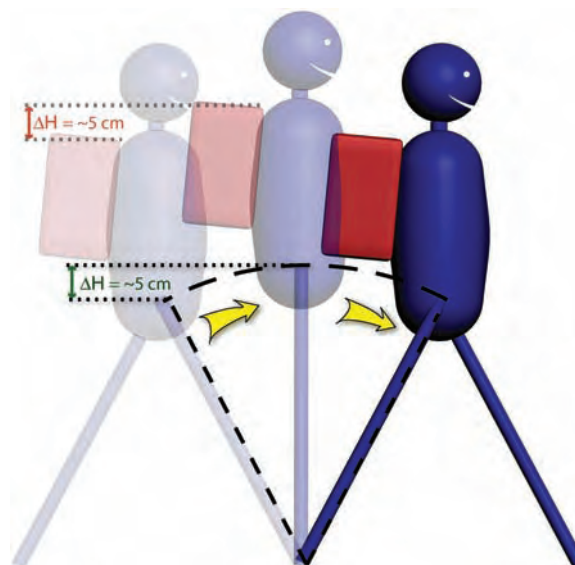

Fig. 1. Humans use an inverted pendulum mode of walking, in which the hip traces out an arc over an extended leg with a vertical excursion (ΔH) of approximately 5 cm. A backpack load rigidly attached to the body would undergo the same vertical excursion. This excursion drives electricity generation.
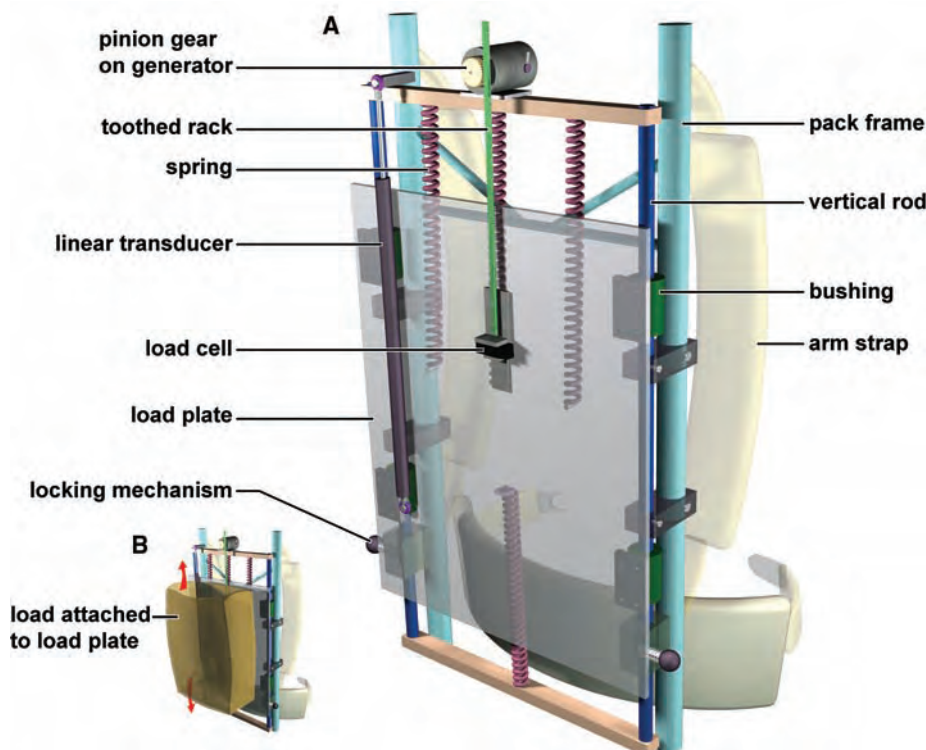

Fig. 2. In the suspended-load backpack, the pack frame is fixed to the body, but the load, mounted on the load plate, is suspended by springs (red) from the frame (blue) (**A**). During walking, the load is free to ride up and down on bushings constrained to vertical rods (**B**) (11). Electricity generation was accomplished by attaching a toothed rack to the load plate, which when moving up and down during walking, meshed with a pinion gear mounted on a geared dc motor, functioning as a generator, rigidly attached to the backpack frame.

erated from shoe devices (10 to 20 mW) (2, 6). The mechanical power removed by the generator (and gears) is the product of the average force exerted on the rack (F_{rack}), the displacement of the load with respect to the pack frame (dl_{rack}) (11), and the step frequency. Mechanical power into the generator increased with speed and load in a similar fashion as electrical power output. Hence, the efficiency of conversion of mechanical energy to electrical energy (that is, electrical power output divided by mechanical power input) was nearly constant (30 to 40%) over this range of speeds and loads.

To power portable devices (or charge batteries), the alternating polarity of the voltage and current (Fig. 3) must be rectified, which the suspended-load backpack can accomplish with little reduction (~5%) in electrical power output (fig. S1) (11). Hence, using circuitry for voltage smoothing, the suspended-load backpack can power multiple devices such as cell phones or GPS receivers, both of which use less than 1 W (11).

If generating electricity while wearing the backpack markedly increased metabolic rate, the device would be of limited use. Indeed, one would expect that because mechanical energy is
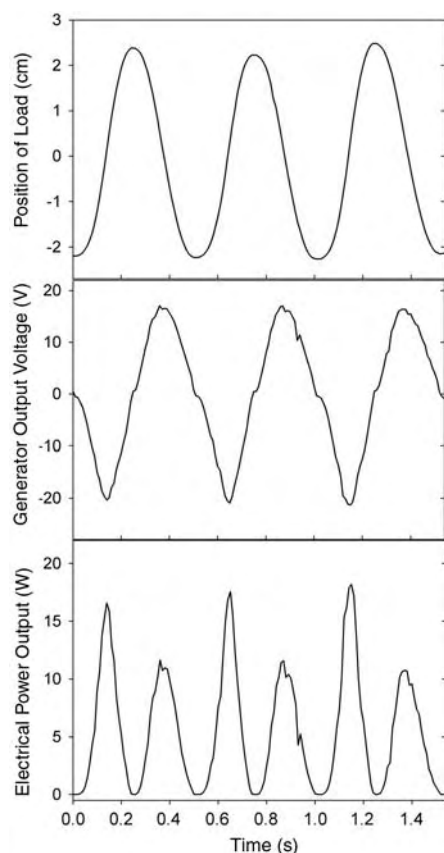
continuously removed from the system by the generator, the muscles would need to perform additional mechanical work during electricity generation in order to replace it. For instance, the mechanical power input to the generator is 12.15 W while walking at 5.6 km hour^{-1} and carrying a 29-kg load (table S3). Because the maximum efficiency of mechanical power production by human muscle is about 25% (14, 15), if the body movement was otherwise the same, one might anticipate a minimum increase of 48.6 W in metabolic power input. We measured the rate of O_2 consumption ($\dot{V}O_2$) and CO_2 production ($\dot{V}CO_2$) of participants walking with the backpack in two configurations: locked (no relative movement, mechanical energy loss, nor electrical energy generation) and unlocked (normal relative movement and electricity generation) in a repeated, paired protocol (11) specifically designed to resolve small differences. We found that the metabolic rate increase (Δmetabolic power input) compared to that with the locked backpack was only about 19.1 W (table S1) (11), which is much less than would be predicted, providing an "apparent efficiency" of mechanical work production of ~63% (table S3).

On the one hand, these results indicate that electricity can be generated metabolically more cheaply than anticipated. But on the other hand, they suggest that there must be some change in gait or loading regime while walking with the unlocked backpack, which causes a reduction of 29.5 W (48.6 minus 19.1 W) or about 3/5 of the metabolic power required for doing work against the generator. Considerable savings in metabolic cost have been previously reported in African women carrying loads on their heads
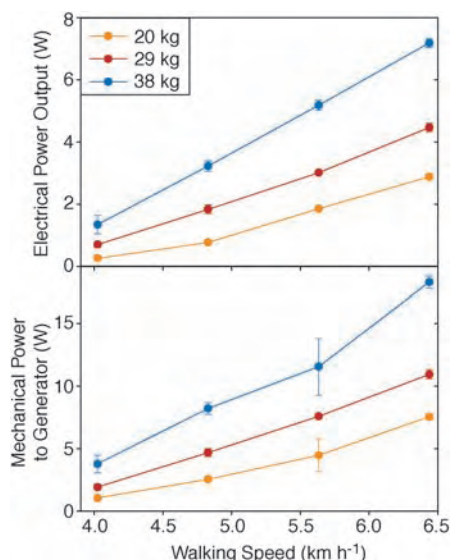
and attributed to more efficient transfer between kinetic and potential energy (16). Although the precise mechanism of compensation in our study remains a mystery, an initial kinematic analysis revealed significant alterations in the biomechanics of walking that could be at the root of the reduction in metabolism. In particular, although there was no change in step frequency, the averaged vertical displacement of the hip during each step was 67.4 mm for the locked condition but only about 55.5 mm, or 11.9 mm less, for the unlocked condition (table S3). Further, there was an 11.8% ($\pm$SE = 1.67%, $n = 4$ participants, $P = 0.008$) reduction in the peak force exerted by the load back onto the person, as well as a change in phasing of this force with respect to the gait cycle (fig. S2). Because these factors will affect the magnitude and time course of forces, as well as the position of the center of mass, they will likely affect the amount of positive work that must be performed during the "double-support phase," a major determinant of the cost of walking (17–21).

Finally, despite the smaller than predicted Δmetabolic power input, individuals may have to carry extra food in order to power electricity generation. This weight, however, is negligible compared to the weight of batteries required to generate the same electrical energy. The specific energy of food (3.9×10^7 J kg^{-1}) (22) is about 100-fold greater than the specific energy of lithium batteries (4.1×10^5 J kg^{-1}) (1) and 35-fold greater than that of zinc-air batteries (1.1×10^6 J kg^{-1}). Given that the "metabolic efficiency of electricity generation" (electricity power output/Δmetabolic power input) is 19.5% (table S3), the extra food to be used for generating electricity would require about 20- and 6.8-fold less weight than lithium and zinc-air batteries, respectively (23). Hence, the longer the expedition, the greater the weight savings (24). Further, the ~12% reduction in peak force exerted on the body by a given load (fig. S2), as well as the potential for using some of the extracted mechanical energy directly for cooling the user (such as through forced air ventilation with a fan or pumping of a coolant) would provide additional ergonomic benefits to the user.

Throughout history, humans have solved many problems by inventing passive devices to enhance the movements made by their muscles (such as springy bamboo poles to carry loads or skis to move through the snow) (25, 26). The suspended-load backpack is another passive device that may help solve a growing problem in the 21st century.

Fig. 3. Position, generator voltage, and electrical power output during walking. In these records, the person is walking with a 38-kg load at 5.6 km hour^{-1}. "Position" refers to the position of the load with respect to the backpack frame, with the midpoint labeled as 0 cm.

Fig. 4. The electrical power output and mechanical power input of the generator as a function of walking speed and load. This graph shows average values of four separate trials for each of 12 conditions for one person. The standard deviation bars are shown but are often smaller than the symbol.

References and Notes
1. J. L. Gonzalez, A. Rubio, F. Moll, *Int. J. Soc. Mater. Eng. Resour.* **10**, 34 (2002).
2. J. Kymissis, C. Kendall, J. Paradiso, N. Gershenfeld, in *IEEE International Conference on Wearable Computing* (IEEE Computer Society Press, Los Alamitos, CA, 1998), pp. 132–139.

3. S. Georgi, www.batteriesdigest.com/id380.htm (accessed June 2005).
4. R. M. Alexander, *J. Exp. Biol.* **160**, 55 (1991).
5. G. A. Cavagna, N. C. Heglund, C. R. Taylor, *Am. J. Physiol.* **233**, R243 (1977).
6. J. Drake, *Wired* **9**, 90 (2001).
7. S. Stanford, R. Pelrine, R. Kornbluh, Q. Pei, in *Proceedings of the 13th International Symposium on Unmanned Untethered Submersible Technology* (Autonomous Undersea Systems Institute, Lee, NH, 2003).
8. T. Starner, J. Paradiso, in *Low Power Electronics Design* (CRC Press, Boca Raton, FL, 2004), p. 45–1.
9. G. A. Cavagna, M. Kaneko, *J. Physiol.* **268**, 647 (1977).
10. S. A. Gard, S. C. Miff, A. D. Kuo, *Hum. Mov. Sci.* **22**, 597 (2004).
11. Supporting material is available on *Science* Online.
12. Because it is a prototype, there has been no attempt to reduce the weight of the backpack—indeed, it is substantially "overdesigned." Further, the 5.6 kg includes the weight of six load cells and one 25-cm-long transducer, each with accompanying brackets and cables, as well as other components that will not be present on a typical pack. In future prototypes, we estimate that the weight will exceed that of a normal backpack by no more than 1 to 1.5 kg.
13. Under high-power conditions (5.6 km hour^{-1} with 20- and 29-kg loads and 4.8 km hour^{-1} with a 38-kg load), power generation on the incline was the same as on the flat. Under low-power conditions (4.8 km hour^{-1} with 20- and 28-kg loads), electricity generation on the incline was actually substantially greater than that on the flat (table S1).
14. R. Margaria, *Biomechanics and Energetics of Muscular Exercise* (Clarendon, Oxford, 1976).
15. R. A. Ferguson *et al.*, *J. Physiol.* **536**, 261 (2001).
16. G. A. Cavagna, P. A. Willems, M. A. Legramandi, N. C. Heglund, *J. Exp. Biol.* **205**, 3413 (2002).
17. A. Grabowski, C. T. Farley, R. Kram, *J. Appl. Physiol.* **98**, 579 (2005).
18. J. M. Donelan, R. Kram, A. D. Kuo, *J. Exp. Biol.* **205**, 3717 (2002).
19. J. M. Donelan, R. Kram, A. D. Kuo, *J. Biomech.* **35**, 117 (2002).
20. J. S. Gottschall, R. Kram, *J. Appl. Physiol.* **94**, 1766 (2003).
21. Because this savings in metabolic energy represents only 6% of the net energetic cost of walking with the backpack (492 W) (table S3) (*17*, *18*), accurate determinations of the position and movements of the center of mass, as well as the direction and magnitude of the ground reaction forces, are essential to discern the mechanism. This will require twin–force-platform single-leg measurements, as well as a complete kinematics and mechanical energy analysis (*19*, *20*). The energy analysis is made more complex because the position of the load with respect to the backpack frame and the amount of energy stored in the backpack springs vary during the gait cycle. Finally, electromyogram measurements are also important to test whether a change in effective muscle moment arms may have caused a change in the volume of activated muscle and hence a change in metabolic cost (*20*, *27*, *28*).
22. K. Schmidt-Nielsen, *Animal Physiology: Adaptation and Environment* (Cambridge Univ. Press, Cambridge, ed. 3, 1988).
23. This assumes that electronic devices are being powered in real time. If there were a power loss of 50% associated with storage (such as in batteries) and re-covery of electrical energy, then these factors would be halved.
24. When not walking, the rack can be disengaged and the generator cranked by hand or by foot. Electrical powers of ~3 W are achievable by hand, and higher wattage can be achieved by using the leg to power it.
25. R. Kram, *J. Appl. Physiol.* **71**, 1119 (1991).
26. A. E. Minetti, *J. Exp. Biol.* **207**, 1265 (2004).
27. A. A. Biewener, C. T. Farley, T. J. Roberts, M. Temaner, *J. Appl. Physiol.* **97**, 2266 (2004).
28. T. M. Griffin, T. J. Roberts, R. Kram, *J. Appl. Physiol.* **95**, 172 (2003).
29. This work was supported by NIH grants AR46125 and AR38404. Some aspects of the project were supported by Office of Naval Research grant N000140310568 and a grant from the University of Pennsylvania Research Foundation. The authors thank Q. Zhang, H. Hofmann, W. Megill, and A. Dunham for helpful discussions; R. Sprague, E. Maxwell, R. Essner, L. Gazit, M. Yuhas, and J. Milligan for helping with the experimentation; and F. Letterio for machining the backpacks.

Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome

Jay Shendure,[1]*† Gregory J. Porreca,[1]*† Nikos B. Reppas,[1] Xiaoxia Lin,[1] John P. McCutcheon,[2,3] Abraham M. Rosenbaum,[1] Michael D. Wang,[1] Kun Zhang,[1] Robi D. Mitra,[2] George M. Church[1]

We describe a DNA sequencing technology in which a commonly available, inexpensive epifluorescence microscope is converted to rapid nonelectrophoretic DNA sequencing automation. We apply this technology to resequence an evolved strain of *Escherichia coli* at less than one error per million consensus bases. A cell-free, mate-paired library provided single DNA molecules that were amplified in parallel to 1-micrometer beads by emulsion polymerase chain reaction. Millions of beads were immobilized in a polyacrylamide gel and subjected to automated cycles of sequencing by ligation and four-color imaging. Cost per base was roughly one-ninth as much as that of conventional sequencing. Our protocols were implemented with off-the-shelf instrumentation and reagents.

The ubiquity and longevity of Sanger sequencing (*1*) are remarkable. Analogous to semiconductors, measures of cost and production have followed exponential trends (*2*). High-throughput centers generate data at a speed of 20 raw bases per instrument-second and a cost of $1.00 per raw kilobase. Nonetheless, optimizations of electrophoretic methods may be reaching their limits. Meeting the challenge of the $1000 human genome requires a paradigm shift in our underlying approach to the DNA polymer (*3*).

Cyclic array methods, an attractive class of alternative technologies, are "multiplex" in that they leverage a single reagent volume to enzymatically manipulate thousands to millions of immobilized DNA features in parallel. Reads are built up over successive cycles of imaging-based data acquisition. Beyond this common thread, these technologies diversify in a panoply of ways: single-molecule versus multimolecule features, ordered versus disordered arrays, sequencing biochemistry, scale of miniaturization, etc. (*3*). Innovative proof-of-concept experiments have been reported, but are generally limited in terms of throughput, feature density, and library complexity (*4–9*). A range of practical and technical hurdles separate these test systems from competing with conventional sequencing on genomic-scale applications.

Our approach to developing a more mature alternative was guided by several considerations. (i) An integrated sequencing pipeline includes library construction, template amplification, and DNA sequencing. We therefore sought compatible protocols that multiplexed each step to an equivalent order of magnitude. (ii) As more genomes are sequenced de novo, demand will likely shift toward genomic resequencing; e.g., to look at variation between individuals. For resequencing, consensus accuracy increases in importance relative to read length because a read need only be long enough to correctly position it on a reference genome. However, a consensus accuracy of 99.99%, i.e., the Bermuda standard, would still result in hundreds of errors in a microbial genome and hundreds of thousands of errors in a mammalian genome. To avoid unacceptable numbers of false-positives, a consensus error rate of 1×10^{-6} is a more reasonable standard for which to aim. (iii) We sought to develop sequencing chemistries compatible with conventional epifluorescence imaging. Diffraction-limited optics with charge-coupled device detection achieves an excellent balance because it not only provides submicrometer resolution and high sensitivity for rapid data acquisition, but is also inexpensive and easily implemented.

[1]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. [2]Department of Genetics, [3]Howard Hughes Medical Institute, Washington University, St. Louis, MO 63110, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: shendure@alumni.princeton.edu (J.S.), gregory_porreca@student.hms.harvard.edu (G.J.P.)

Conventional shotgun libraries are constructed by cloning fragmented genomic DNA of a defined size range into an *Escherichia coli* vector. Sequencing reads derived from opposite ends of each fragment are termed "mate-pairs." To avoid bottlenecks imposed by *E. coli* transformation, we developed a multiplexed, cell-free library construction protocol. Our strategy (Fig. 1A) uses a type IIs restriction endonuclease to bring sequences separated on the genome by ~1 kb into proximity. Each ~135–base pair (bp) library molecule contains two mate-paired 17- to 18-bp tags of unique genomic sequence, flanked and separated by universal sequences that are complementary to amplification or sequencing primers used in subsequent steps. The in vitro protocol (Note S1) results in a library with a complexity of ~1 million unique, mate-paired species.

Conventionally, template amplification has been performed by bacterial colonies that must be individually picked. Polymerase colony, or polony, technologies perform multiplex amplification while maintaining spatial clustering of identical amplicons (*10*). These include in situ polonies (*11*), in situ rolling circle amplification (RCA) (*12*), bridge polymerase chain reaction (PCR) (*13*), picotiter PCR (*9*), and emulsion PCR (*14*). In emulsion PCR (ePCR), a water-in-oil emulsion permits millions of noninteracting amplifications within a milliliter-scale volume (*15–17*). Amplification products of individual compartments are captured via inclusion of 1-μm paramagnetic beads bearing one of the PCR primers (*14*). Any single bead bears thousands of single-stranded copies of the same PCR product, whereas different beads bear the products of different compartmentalized PCR reactions (Fig. 1B). The beads generated by ePCR have highly desirable characteristics: high signal density, geometric uniformity, strong feature separation, and a size that is small but still resolvable by inexpensive optics.

Provided that the template molecules are sufficiently short (fig. S1), an optimized version of the ePCR protocol described by Dressman *et al.* (*14*) robustly and reproducibly amplifies our complex libraries (Note S2). In practice, ePCR yields empty, clonal, and nonclonal beads, which arise from emulsion compartments that initially have zero, one, or multiple template molecules, respectively. Increasing template concentration in an ePCR reaction boosts the fraction of amplified beads at the cost of greater nonclonality (*14*). To generate populations in which a high fraction of beads was both amplified and clonal, we developed a hybridization-based in vitro enrichment method (Fig. 1C). The protocol is capable of a fivefold enrichment of amplified beads (Note S3).

Iterative interrogation of ePCR beads (Fig. 1D) requires immobilization in a format compatible with enzymatic manipulation and epifluorescence imaging. We found that a simple acrylamide-based gel system developed for in situ polonies (*6*) was easily applied to ePCR beads, resulting in a ~1.5-cm² array of disordered, monolayered, immobilized beads (Note S4, Fig. 2A).

With few exceptions (*18*), sequencing biochemistries rely on the discriminatory capacities of polymerases and ligases (*1, 6, 8, 19–22*). We evaluated a variety of sequencing protocols in our system. A four-color sequencing by ligation scheme ("degenerate ligation") yielded the most promising results (Fig. 2, B and C). A detailed graphical description of this method is shown in fig. S7. We begin by hybridizing an "anchor primer" to one of four positions (immediately 5′ or 3′ to one of the two tags). We then perform an enyzmatic ligation reaction of the anchor primer to a population of degenerate nonamers that are labeled with fluorescent dyes. At any given cycle, the population of nonamers that is used is structured such that the identity of one of its positions is correlated with the identity of the fluorophore attached to that nonamer. To the extent that the ligase discriminates for complementarity at that queried position, the fluorescent signal allows us to infer the identity of that base (Fig. 2, B and C). After performing the ligation and four-color imaging, the anchor primer:nonamer complexes are stripped and a new cycle is begun. With T4 DNA ligase, we can obtain accurate sequence when the query position is as far as six bases from the ligation junction while ligating in the 5′→3′ direction, and seven bases from the ligation junction in the 3′→5′ direction. This allows us to access 13 bp per tag (a hexamer and heptamer separated by a 4- to 5-bp gap) and 26 bp per amplicon (2 tags × 13 bp) (fig. S7).

Although the sequencing method presented here can be performed manually, we benefited from fully automating the procedure (fig. S3). Our integrated liquid-handling and microscopy setup can be replicated with off-the-shelf components at a cost of about $140,000. A detailed description of instrumentation and software is provided in Notes S5 and S7.

As a genomic-scale challenge, we sought a microbial genome that was expected, relative to a reference sequence, to contain a modest number of both expected and unexpected differences.
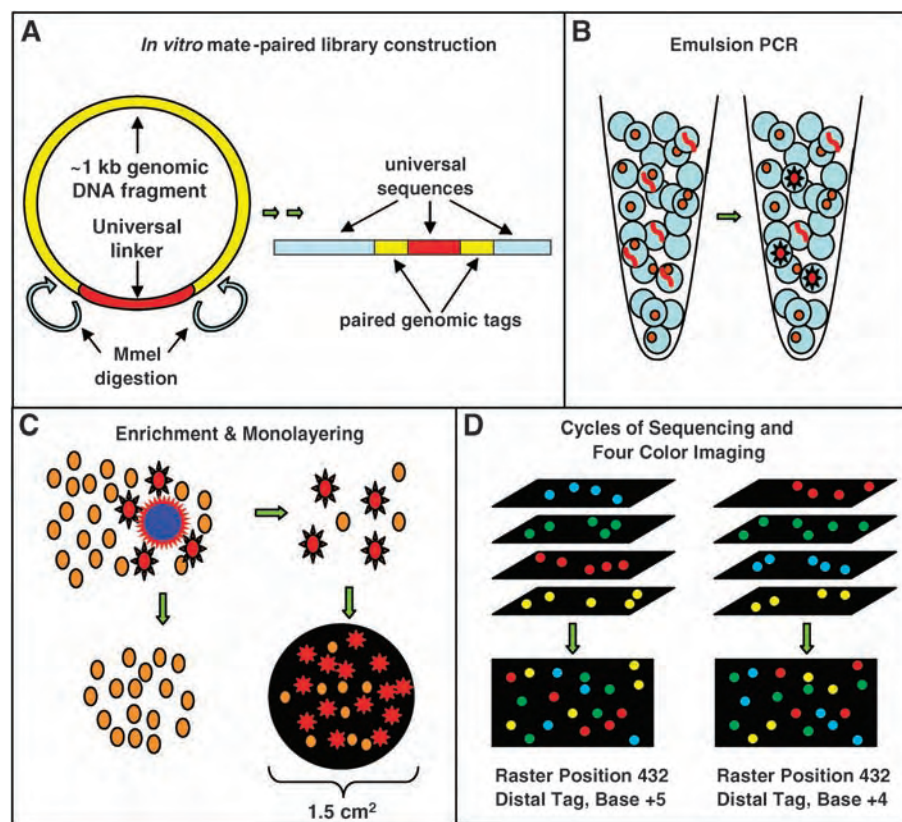

Fig. 1. A multiplex approach to genome sequencing. (**A**) Sheared, size-selected genomic fragments (yellow) are circularized with a linker (red) bearing Mme I recognition sites (Note S1). Subsequent steps, which include a rolling circle amplification, yield the 134- to 136-bp mate-paired library molecules shown at right. (**B**) ePCR (*14*) yields clonal template amplification on 1-μm beads (Note S2). (**C**) Hybridization to nonmagnetic, low-density "capture beads" (dark blue) permits enrichment of the amplified fraction (red) of magnetic ePCR beads by centrifugation (Note S3). Beads are immobilized and mounted in a flowcell for automated sequencing (Note S4). (**D**) At each sequencing cycle, four-color imaging is performed across several hundred raster positions to determine the sequence of each amplified bead at a specific position in one of the tags. The structure of each sequencing cycle is discussed in the text, Note S6, and fig. S7.

We selected a derivative of *E. coli* MG1655, engineered for deficiencies in tryptophan biosynthesis and evolved for ~200 generations under conditions of syntrophic symbiosis via coculture with a tyrosine biosynthesis–deficient strain (*23*). Specific phenotypes emerged during the laboratory evolution, leading to the expectation of genetic changes in addition to intentionally engineered differences.

An in vitro mate-paired library was constructed from genomic DNA derived from a single clone of the evolved Trp$_v^-$ strain. To sequence this library, we performed successive instrument runs with progressively higher bead densities. In an experiment ultimately yielding 30.1 Mb of sequence, 26 cycles of sequencing were performed on an array containing amplified, enriched ePCR beads. At each cycle, data were acquired for four wavelengths at $20\times$ optical magnification by rastering across each of 516 fields of view on the array (Fig. 1D). A detailed description of the structure of each sequencing cycle is provided in Note S6. In total, 54,696 images (14 bit, 1000×1000) were collected. Cycle times averaged 135 min per base (~90 min for reactions and ~45 min for imaging), for a total of ~60 hours per instrument run.

Image processing and base calling algorithms are detailed in Note S7. In brief, all images taken at a given raster position were aligned. Two additional image sets were acquired: brightfield images to robustly identify bead locations (Fig. 2A) and fluorescent primer images to identify amplified beads. Our algorithms detected 14 million objects within the set of brightfield images. On the basis of size, fluorescence, and overall signal coherence over the course of the sequencing run, we determined 1.6 million to be well-amplified, clonal beads (~11%). For each cycle, mean intensities for amplified beads were extracted and normalized to a 4D unit vector (Fig. 2, B and C). The Euclidean distance of the unit vector for a given raw base call to the median centroid of the nearest cluster serves as a natural metric of the quality of that call.

The reference genome consisted of the *E. coli* MG1655 genome (GenBank accession code U00096.2) appended with sequences corresponding to the *cat* gene and the lambda Red prophage, which had been engineered into the sequenced strain to replace the *trp* and *bio* operons, respectively. To systematically assess our power to detect single-base substitutions, we introduced a set of 100 random single-nucleotide changes into the reference sequence at randomly selected positions ("mock SNCs") (Table 1).

An algorithm was developed to place the discontinuous reads onto the reference sequence (Note S7). The matching criteria required the paired tags to be appropriately oriented and located within 700 to 1200 bp of one another, allowing for substitutions if exact matches

were not found. Of the 1.6 million reads, we were able to confidently place ~1.16 million (~72%) to specific locations on the reference genome, resulting in ~30.1 million bases of resequencing data at a median raw accuracy of 99.7%. At this stage of the analysis, the data were combined with reads from a previous instrument run that contributed an additional ~18.1 million bases of equivalent quality (Fig. 2D). In this latter experiment, ~1.8 million reads were generated from ~7.6 million objects (~24%), of which ~0.8 million were confidently placed (~40%).

High-confidence consensus calls were determined for 70.5% of the *E. coli* genome for which sufficient and consistent coverage was available (3,289,465 bp; generally positions with ~4× or greater coverage). There were six positions within this set that did not agree with the reference sequence, and thus were targeted for confirmation by Sanger sequencing. All six were correct, although in one case we detected the edge of an 8-bp deletion rather than a substitution (Table 2). Three of these six mutations represent heterogeneities in lambda Red or MG1655, or errors in the
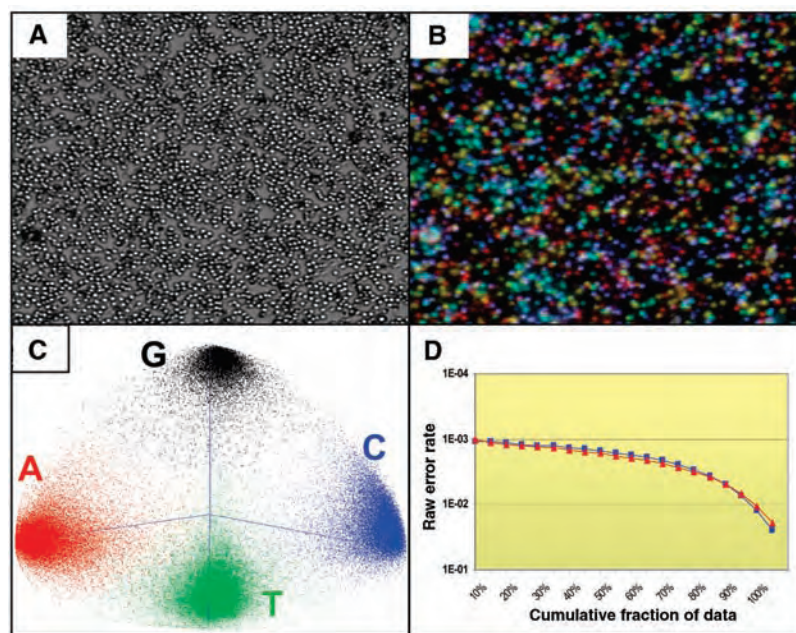

Fig. 2. Raw data acquisition and base calling. (**A**) Brightfield images (area shown corresponds to 0.01% of the total gel area) facilitate object segmentation by simple thresholding, allowing resolution even when multiple 1-μm beads are in contact. (**B**) False-color depiction of four fluorescence images acquired at this location from a single ligation cycle. A, gold; G, red; C, light blue; T, purple. (**C**) Four-color data from each cycle can be visualized in tetrahedral space, where each point represents a single bead, and the four clusters correspond to the four possible base calls. Shown is the sequencing data from position (-1) of the proximal tag of a complex *E. coli*–derived library. (**D**) Cumulative distribution of raw error as a function of rank-ordered quality for two independent experiments (red triangles, 18.1-Mb run; blue squares, 30.1-Mb run). The *x* axis indicates percentile bins of beads, sorted on the basis of a confidence metric. The *y* axis (logarithmic scale) indicates the raw base-calling accuracy of each cumulative bin. Equivalent Phred scores are Q20 = 1×10^{-2}, Q30 = 1×10^{-3} {Phred score = $-10[\log_{10}(\text{raw per-base error})]$}. Cumulative distribution of raw error with sequencing by ligation cycles considered independently is shown in fig. S8.

Table 1. Genome Coverage and SNC prediction. Bases with consistent consensus coverage were used to make mutation predictions. To assess power, the outcome of consensus calling for the mock SNC positions with various levels of coverage was determined. Data from two independent sets of mock SNCs are shown. "86 of 87," for example, means that 87 of the 100 mock SNCs were present in the sequence that was covered with 1× or more reads, and 86 of these were called correctly.

Coverage	Percent of genome	Correctly called mock substitutions
1× or greater	91.4%	86 of 87
		88 of 90
2× or greater	83.3%	78 of 78
		75 of 76
3× or greater	74.9%	67 of 67
		68 of 68
4× or greater	66.9%	58 of 58
		62 of 62

reference sequence; three were only present in the evolved variant (Table 2). Of the 100 mock SNCs, 53 were at positions called with high confidence. All of these were correctly called as substitutions of the expected nucleotide (59 of 59 on a second set of mock SNCs). The absence of substitution errors in ~3.3 Mb of reference sequence positions called with high confidence suggests that we are achieving consensus accuracies sufficient for resequencing applications. Percentage of the genome covered and mock SNC discovery at various levels of coverage are shown in Table 1.

Despite 10× coverage in terms of raw base pairs, only ~91.4% of the genome had at least

1× coverage (fig. S4). Substantial fluctuations in coverage were observed owing to the stochasticity of the RCA step of library construction. We are currently generating libraries that are more complex and more evenly distributed.

A Gaussoid distribution of distances between mate-paired tags was observed, consistent with the size selection during library construction (Fig. 3, A and B). Notably, the helical pitch of DNA (~10.6 bp per turn) is evident in the local statistics of ~1 million circularization events (Fig. 3B). As a function of the number of bases sequenced, we generated over an order of magnitude more mate-pairing data points than an equivalent amount of conventional sequenc-

ing. To detect genomic rearrangements, we mined the unplaced mate-pairs for consistent links between genomic regions that did not fall within the expected distance constraints. In addition to detecting the expected replacements of the *trp* and *bio* operons with *cat* and lambda Red prophage (Fig. 3D), we detected and confirmed the absence of a 776-bp IS1 transposon (Fig. 3C), a previously described heterogeneity in MG1655 strains (*24*). We also detected and confirmed a ~1.8-kb region that was heterogeneously inverted in the genomic DNA used to construct the library (Fig. 3E), owing to activity of *pin* on the invertible P region (*25*).

We observe error rates of ~0.001 for the better half of our raw base calls (Fig. 2D). Although high consensus accuracies are still achieved with relatively low coverage, our best raw accuracies are notably one to two orders of magnitude less accurate than most raw bases in a conventional Sanger sequencing trace. The PCR amplifications before sequencing are potentially introducing errors at a rate that imposes a ceiling on the accuracies achievable by the sequencing method itself. One potential solution is to create a library directly from the genomic material to be sequenced, such that the library molecules are linear RCA amplicons. Such concatemers, where each copy is independently derived from the original template, would theoretically provide a form of error correction during ePCR.

Our algorithms were focused on detection of point substitutions and rearrangements. Increasing read lengths, currently totaling only 26 bp per amplicon, will be critical to detecting a wider spectrum of mutation. A higher fidelity ligase (*20*) or sequential nonamer ligations (*20*, *21*) may enable completion of each 17- to 18-bp tag. Eco P15 I, which generates ~27-bp tags, would allow even longer read lengths while retaining the same mate-pairing scheme (*26*).

We estimate a cost of $0.11 per raw kilobase of sequence generated (Note S8), roughly one-ninth as much as the best costs for electrophoretic sequencing. Raw data in all sequencing methods are generally combined to form a consensus. Even though costs are generally defined in terms of raw bases, the critical metric to compare technologies is consensus accuracy for a given cost. There is thus a need to devise appropriate cost metrics for specific levels of consensus accuracy.

If library construction costs are not included, the estimated cost drops to $0.08 per raw kilobase. Higher densities of amplified beads are expected to boost the number of bases sequenced per experiment. While imaging, data were collected at a rate of ~400 bp/s. Although enzymatic steps slowed our overall throughput to ~140 bp/s, a dual flowcell instrument (such that the microscope is always imaging) will allow us to achieve continuous data acquisition. Enzymatic reagents, which dominate our cost equation, can be produced in-house at a fraction of the commercial price.
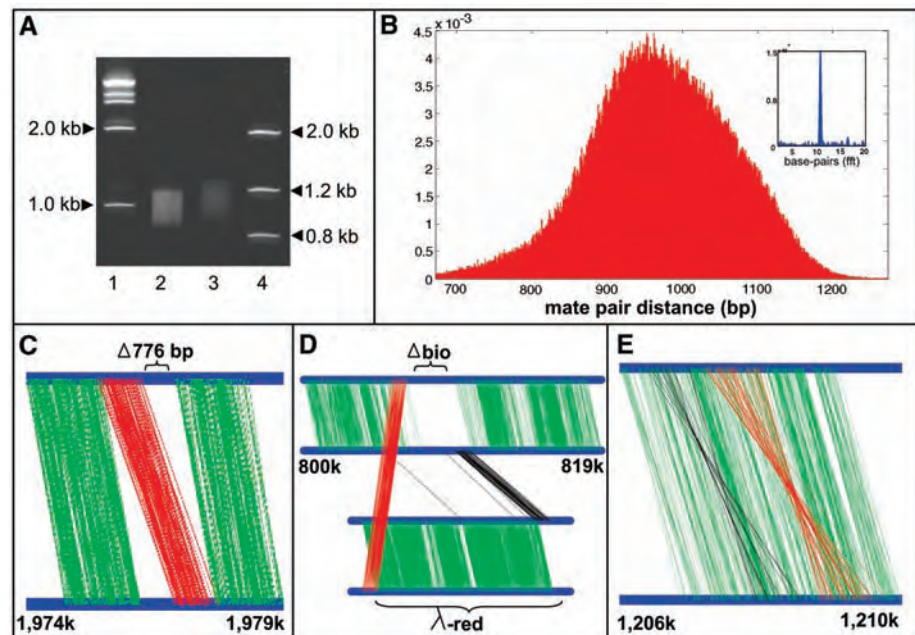

Fig. 3. Mate-paired tags and rearrangement discovery. (**A**) Diagnostic 6% polyacrylamide gel of the sheared, size-selected genomic DNA from which the library was constructed. Lanes 1 and 4 are molecular size markers. Lane 2 represents the material used in the library sequenced to generate the paired-tag mappings in (B), and lane 3 represents genomic DNA for a different library. (**B**) Histogram of distances between ~1 million mapped mate-pair sequences. The probability of circularization favors integrals of the helical pitch of DNA, such that the Fourier transform of the distribution (**inset**) yields a peak at 10.6 bp (*27*) (**C** to **E**). Consistent, aberrant mapping of unplaced mate-pairs to distal sequences revealed information about underlying rearrangements. Top and bottom blue bars indicate genomic positions for proximal and distal tags, respectively. Green connections indicate mate-pairings that fall within expected distance constraints, whereas red and black connections indicate aberrant connections (red indicates connections between the same strand, and black, connections between opposite strands). (**C**) Detection of a 776-bp deletion in the *flhD* promoter (*24*). (**D**) Detection of the replacement of the *bio* locus with the lambda red construct. (**E**) Detection of the P-region inversion (*25*). Detection of the inversion on a background of normally mate-paired reads indicates that the inversion is heterogeneously present.

Table 2. Polymorphism discovery. Predictions for mutated positions were tested and verified as correct by Sanger sequencing. We found three mutations unique to the evolved strain—two in *ompF*, a porin, and one in *lrp*, a global regulator.

Position	Type	Gene	Context	Confirmation	Comments
986,328	T → G	*ompF*	−10 region	Yes	Evolved strain only
931,955	8-bp deletion	*lrp*	Frameshift	Yes	Evolved strain only
985,791	T → G	*ompF*	Glu → Ala	Yes	Evolved strain only
1,976,527 − 1,977,302	776-bp deletion	*flhD*	Promoter	Yes	MG1655 heterogeneity
3,957,960	C → T	*ppiC*	5′ UTR	Yes	MG1655 heterogeneity
λ-red, 3274	T → C	ORF61	Lys → Gly	Yes	λ-red heterogeneity
λ-red, 9846	T → C	*cI*	Glu → Glu	Yes	λ-red heterogeneity

We demonstrate low costs of sequencing, mate-paired reads, high multiplicities, and high consensus accuracies. These enable applications including BAC (bacterial artificial chromosome) and bacterial genome resequencing, as well as SAGE (serial analysis of gene expression) tag and barcode sequencing. Simulations suggest that the current mate-paired libraries are compatible with human genome resequencing, provided that the read length can be increased to cover the full 17- to 18-bp tag (fig. S5).

What are the limits of this approach? As many as 1 billion 1-μm beads can potentially be fit in the area of a standard microscope slide (fig. S6). We achieve raw data acquisition rates of ~400 bp/s, more than an order of magnitude faster than conventional sequencing. From another point of view, we collected ~786 gigabits of image data from which we gleaned only ~60 megabits of sequence. This sparsity—one useful bit of information per 10,000 bits collected—is a ripe avenue for improvement. The natural limit of this direction is single-pixel sequencing, in which the commonplace analogy between bytes and bases will be at its most manifest.

References and Notes

1. F. Sanger et al., Nature 265, 687 (1977).
2. F. S. Collins, M. Morgan, A. Patrinos, Science 300, 286 (2003).
3. J. Shendure et al., Nat. Rev. Genet. 5, 335 (2004).
4. I. Braslavsky, B. Hebert, E. Kartalov, S. R. Quake, Proc. Natl. Acad. Sci. U.S.A. 100, 3960 (2003).
5. T. S. Seo et al., Proc. Natl. Acad. Sci. U.S.A. 102, 5926 (2005).
6. R. D. Mitra, J. Shendure, J. Olejnik, O. Edyta Krzymanska, G. M. Church, Anal. Biochem. 320, 55 (2003).
7. M. J. Levene et al., Science 299, 682 (2003).
8. M. Ronaghi, S. Karamohamed, B. Pettersson, M. Uhlen, P. Nyren, Anal. Biochem. 242, 84 (1996).
9. J. H. Leamon et al., Electrophoresis 24, 3769 (2003).
10. http://arep.med.harvard.edu/Polonator/Plone.htm
11. R. D. Mitra, G. M. Church, Nucleic Acids Res. 27, e34 (1999).
12. P. M. Lizardi et al., Nat. Genet. 19, 225 (1998).
13. C. P. Adams, S. J. Kron, U.S. Patent 5,641,658 (1997).
14. D. Dressman, H. Yan, G. Traverso, K. W. Kinzler, B. Vogelstein, Proc. Natl. Acad. Sci. U.S.A. 100, 8817 (2003).
15. D. S. Tawfik, A. D. Griffiths, Nat. Biotechnol. 16, 652 (1998).
16. F. J. Ghadessy, J. L. Ong, P. Holliger, Proc. Natl. Acad. Sci. U.S.A. 98, 4552 (2001).
17. M. Nakano et al., J. Biotechnol. 102, 117 (2003).
18. A. M. Maxam, W. Gilbert, Proc. Natl. Acad. Sci. U.S.A. 74, 560 (1977).
19. F. Barany, Proc. Natl. Acad. Sci. U.S.A. 88, 189 (1991).
20. J. N. Housby, E. M. Southern, Nucleic Acids Res. 26, 4259 (1998).
21. S. C. Macevicz, U.S. Patent 5,750,341 (1998).
22. S. Brenner et al., Nat. Biotechnol. 18, 630 (2000).
23. N. B. Reppas, X. Lin, in preparation.
24. C. S. Barker, B. M. Pruss, P. Matsumura, J. Bacteriol. 186, 7529 (2004).
25. R. H. Plasterk, P. van de Putte, EMBO J. 4, 237 (1985).
26. M. Mucke, S. Reich, E. Moncke-Buchner, M. Reuter, D. H. Kruger, J. Mol. Biol. 312, 687 (2001).
27. D. Shore, R. L. Baldwin, J. Mol. Biol. 170, 957 (1983).
28. For advice, encouragement, and technical assistance, we are deeply indebted to J. Zhu, S. Douglas, J. Chou, J. Aach, M. Nikku, A. Lee, N. Novikov, and M. Wright (Church Lab); A. Blanchard, G. Costa, H. Ebling, J. Ichikawa, J. Malek, P. McEwan, K. McKernan, A. Sheridan, and D. Smith (Agencourt); S. Skiena (SUNY–Stony Brook) C. Felts (RPI); R. Fincher (Alcott); D. Focht (Bioptechs); and M. Hotfelder and J. Feng (Washington University). We thank B. Vogelstein, J. Edwards, and their groups for assistance with emulsion PCR. This work was supported by the National Human Genome Research Institute–Centers of Excellence in Genomic Science and U.S. Department of Energy–Genomes to Life grants.

PUMA Couples the Nuclear and Cytoplasmic Proapoptotic Function of p53

Jerry E. Chipuk,[1]* Lisa Bouchier-Hayes,[1] Tomomi Kuwana,[1,2] Donald D. Newmeyer,[1] Douglas R. Green[1]*†

The Trp53 tumor suppressor gene product (p53) functions in the nucleus to regulate proapoptotic genes, whereas cytoplasmic p53 directly activates proapoptotic Bcl-2 proteins to permeabilize mitochondria and initiate apoptosis. Here, we demonstrate that a tripartite nexus between Bcl-xL, cytoplasmic p53, and PUMA coordinates these distinct p53 functions. After genotoxic stress, Bcl-xL sequestered cytoplasmic p53. Nuclear p53 caused expression of PUMA, which then displaced p53 from Bcl-xL, allowing p53 to induce mitochondrial permeabilization. Mutant Bcl-xL that bound p53, but not PUMA, rendered cells resistant to p53-induced apoptosis irrespective of PUMA expression. Thus, PUMA couples the nuclear and cytoplasmic proapoptotic functions of p53.
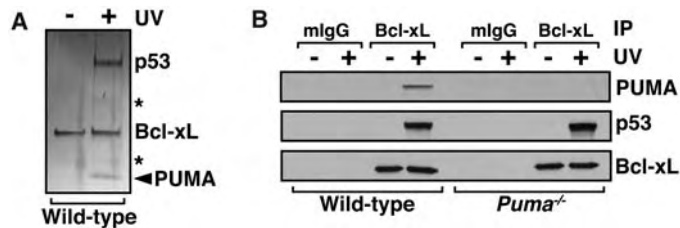
The antineoplastic function of p53 occurs primarily through the induction of apoptosis (1). p53 undergoes posttranslational modification in response to oncogene-activated signaling pathways or to genotoxic stress; this allows stabilization of p53, which accumulates in the nucleus and regulates target gene expression.

Numerous genes are regulated by p53, such as those encoding death receptors [for example, FAS (CD95)] and proapoptotic Bcl-2 proteins (for example, BAX, BID, Noxa, and PUMA)

(2–7). In parallel, p53 also accumulates in the cytoplasm, where it directly activates the proapoptotic protein BAX to promote mitochondrial outer-membrane permeabilization (MOMP) (8–10). Once MOMP occurs, proapoptogenic factors (for example, cytochrome c) are released from mitochondria, caspases are activated, and apoptosis rapidly ensues (11). Thus, p53 possesses a proapoptotic function that is independent of its transcriptional activity (12–15).

If p53 directly engages MOMP in cooperation with BAX, no further requirement for p53-dependent transcriptional regulation of additional proapoptotic Bcl-2 proteins would be expected. Nevertheless, PUMA (p53–up-regulated modifier of apoptosis), a proapoptotic BH3-only protein, is a direct transcriptional target of p53. Furthermore, mice deficient in Puma are resistant to p53-dependent, DNA damage–induced apoptosis even though p53 is stabilized and accumulates in the cytoplasm (6, 16–18). A better understanding of the distinct nuclear and cytoplasmic proapoptotic functions of p53 may reveal strategies for the prevention and treatment of cancer.

[1]Division of Cellular Immunology, La Jolla Institute for Allergy and Immunology, 10355 Science Center Drive, San Diego, CA 92121, USA. [2]University of Iowa, Carver College of Medicine, Department of Pathology, Iowa City, IA 52242, USA.

*Present address: Department of Immunology, St. Jude Children's Research Hospital, 332 North Lauderdale Street, Memphis, TN 38105, USA.
†To whom correspondence should be addressed. E-mail: dgreen5240@aol.com

Fig. 1. DNA damage–induced p53·Bcl-xL and PUMA·Bcl-xL complexes. (A) Proteins from cytosolic extracts prepared from wild-type or Puma$^{-/-}$ MEFs treated with 5 mJ/cm² UV were immunoprecipitated with an agarose-conjugated antibody to Bcl-xL, eluted, subjected to SDS-PAGE, and visualized by silver staining. Bands were excised and subjected to tryptic digestion and mass spectrometry. The asterisk (*) indicates a fragment of Bcl-xL or p53. (B) Cytosolic extracts were treated as in (A), but protein complexes were analyzed by Western blot. mIgG (mouse immunoglobulin G) is a control antibody.

In addition to the ability of cytoplasmic p53 to directly activate BAX, cytoplasmic p53 also interacts with Bcl-xL (8, 9). This interaction has been suggested to be either inhibitory to MOMP induced by p53 and BAX (9); or an initiating apoptotic signal if p53 inactivates the antiapoptotic function of Bcl-xL (8). The association between cytoplasmic p53 and Bcl-xL could prevent p53 from activating BAX; therefore, we screened for potential regulators of the p53·Bcl-xL complex. We prepared cytoplasmic extracts from mouse embryonic fibroblasts (MEFs) after treatment with ultraviolet (UV) light. Several proteins were immunoprecipitated in a complex with Bcl-xL and were detectable after SDS–polyacrylamide gel electrophoresis (SDS-PAGE) and visualization by silver stain (Fig. 1A). Several bands were identified as p53 or fragments of Bcl-xL (fig. S1A), but a major band with a relative molecular mass of 23 kD coprecipitated with Bcl-xL and was excised for tryptic digestion and analyzed by liquid chromatography and tandem mass spectrometry. Three tryptic peptides from the excised band were derived from murine PUMA (fig. S1A) (6). This band was absent when we precipitated Bcl-xL from UV-treated E1A- and H-rasG12V–expressing MEFs that did not express *Puma* (fig. S1B). Protein immunoblotting also confirmed association of PUMA and p53 with Bcl-xL (Fig. 1B). PUMA appeared to be the major regulator of the p53·Bcl-xL complex after UV treatment because no free cytoplasmic p53 was observed in *Puma*$^{-/-}$ cytosol (fig. S1, C and D).

We next determined the kinetics of the p53·Bcl-xL complex after UV-induced DNA damage in the absence or presence of *PUMA* expression (fig. S2C). Bcl-xL was immunoprecipitated at the indicated time points after UV treatment, and p53 association was determined. In the presence of *PUMA* expression (HCT116$^{p21-/-PUMA+/+}$, referred to as *PUMA*$^{+/+}$), the amount of p53 associated with Bcl-xL decreased after UV treatment, and this temporally correlated with the induction of apoptosis (fig. S2C). In contrast, the amount of p53 associated with Bcl-xL remained unchanged after treatment in HCT116$^{p21-/-PUMA-/-}$ (*PUMA*$^{-/-}$) cells, and they were resistant to apoptosis (fig. S2, A and C). Furthermore, UV-induced apoptosis in the HCT116 model system required *BAX* expression (fig. S2D).

To examine the possible effects of PUMA on the stability of the p53·Bcl-xL association, we added a complex of recombinant Bcl-xL and purified cytoplasmic p53 to cytosol samples from UV-irradiated HCT116 cells (Fig. 2A). Cytosol from UV-treated HCT116$^{p21-/-PUMA+/+}$ cells disrupted the complex, such that less p53 was associated with Bcl-xL, and Bcl-xL associated with PUMA from the cytosol. In contrast, cytosol from UV-treated HCT116$^{p21-/-PUMA-/-}$ cells allowed for p53 to remain associated with

Bcl-xL. The p53·Bcl-xL complex was disrupted by incubation with a PUMA BH3 domain peptide or full-length PUMA protein but not by other proapoptotic Bcl-2 protein BH3 domain peptides (Noxa, BIM, BID, BMF, BAD, bNIP3, BIK, HRK, BAX, or BAK) or proteins (BID or BAX) (Fig. 2B). The relative dissociation constants for p53 ($K_D = 164 \pm 54$ nM) and PUMA ($K_D = 10 \pm 4$ nM) from Bcl-xL

were determined by surface plasmon resonance (BIAcore) (Fig. 2C). The p53 released from Bcl-xL by PUMA still activated BAX and induced mitochondrial cytochrome c release (Fig. 2D). Thus, PUMA may function to liberate p53 from Bcl-xL.

Because Bcl-xL expression remains constant after UV treatment in HCT116 cells (Fig. 2A and figs. S2C and S3A), we hypothesized that if
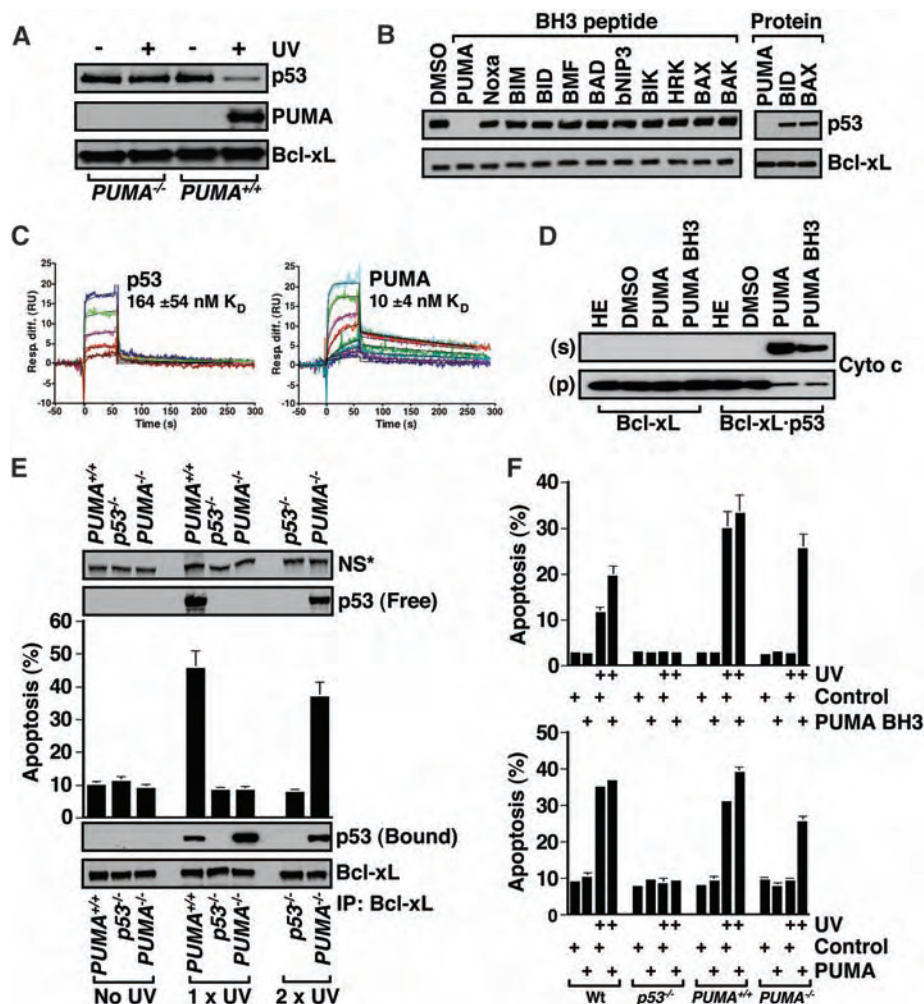

Fig. 2. PUMA-induced release of p53 from Bcl-xL and consequent apoptosis. (**A** and **B**) Recombinant $^{6\times His}$Bcl-xL·p53 complexes (10 nM) were added to cytosol from 100 mJ/cm² UV-treated *PUMA*$^{+/+}$ or *PUMA*$^{-/-}$ cells (A) or 10 μM BH3 peptides of the indicated proapoptotic Bcl-2 proteins or 50 nM proapoptotic Bcl-2 proteins (B). $^{6\times His}$Bcl-xL was then isolated by affinity chromatography, and associated p53 and PUMA were detected by Western blot. (**C**) Dissociation constants (K_D) for p53 and PUMA from immobilized Bcl-xLΔC were determined by surface plasmon resonance. The concentration range of p53 and PUMA analyzed was 0.02 to 1.28 μM and 0.01 to 1.25 μM, respectively. Data are presented as a real-time graph of response units (RU) against time. (**D**) Cytochrome c release from isolated mitochondria. Recombinant 45 nM $^{6\times His}$Bcl-xL·p53 complexes (or $^{6\times His}$Bcl-xL alone) were added to either 45 nM PUMA protein or 10 μM PUMA BH3 peptide, and supernatant containing p53 was then added to mitochondria in the presence of 120 nM recombinant, monomeric full-length BAX. Cytochrome c release was assayed by Western blot by its change in localization from the mitochondrial pellet (p) to the supernatant (s). (**E**) Apoptosis of HCT116 cells. Indicated HCT116 cell lines were treated with two 50 mJ/cm² doses of UV, 72 hours apart, then incubated for an additional 24 hours before harvesting ("2 × UV"); or they were treated once with 50 mJ/cm² UV and incubated for 24 hours before harvesting ("1 × UV"). Cells were analyzed for apoptosis, and cytoplasmic lysates were isolated. Proteins from cytoplasmic lysates were immunoprecipitated (IP) with an antibody to Bcl-xL and evaluated by Western blot. Equal loading is shown by a nonspecific (NS*) band. (**F**) Indicated HCT116 cell lines were transfected with the PUMA BH3 domain peptide or control peptide (top) or pCMV.Flag-PUMA or control vector (bottom), cultured overnight, treated with 75 mJ/cm² UV, and analyzed 24 hours later for apoptosis. Error bars represent ±SD.

large amounts of stabilized p53 accumulate, a fraction of the cytosolic p53 might remain unassociated with Bcl-xL, eliminating the requirement for PUMA to release cytosolic p53 from Bcl-xL. After exposure of HCT116[p21−/−PUMA+/+] and HCT116[p21−/−PUMA−/−] cells to a single dose of UV, cytosolic p53 coimmunoprecipitated with Bcl-xL. In cells expressing PUMA, cytosolic p53 was also found unassociated with Bcl-xL, and this correlated with apoptosis (Fig. 2E). The Bcl-xL·p53 complex was stable for at least 72 hours after UV treatment in HCT116[p21−/−PUMA−/−] (fig. S3B). After a second UV treatment, cytosolic p53 not associated with Bcl-xL was observed in cells lacking PUMA, and these cells underwent apoptosis (Fig. 2E). As a control, HCT116[p53−/−] cells were resistant to two doses of UV. One explanation is that after a second UV treatment, the p53-sequestering activity of Bcl-xL was saturated, thus allowing p53 to engage MOMP. The requirement for *PUMA* expression in p53-dependent, UV-induced apoptosis was also overcome by introduction of the PUMA BH3 domain peptide or transient expression of *PUMA* (Fig. 2F) in HCT116[p21−/−PUMA−/−] cells. Introduction of PUMA BH3 domain peptide or transient expression of PUMA did not induce apoptosis in any of the cell lines, but sensitized the HCT116[p21−/−PUMA−/−] cells to UV-induced death. In contrast, neither PUMA BH3 domain peptide nor PUMA expression sensitized HCT116[p53−/−] to UV-induced apoptosis. Therefore, PUMA is not sufficient for apoptosis or sensitization to UV-induced apoptosis in the absence of p53 (Fig. 2F). Consistent with this, neither PUMA protein nor the BH3 domain peptide directly permeabilized mouse liver mitochondria or induced apoptosis (fig. S4, A to C).

We also tested other BH3 domain peptides for sensitization to UV-induced apoptosis. The BAD and BIM BH3 domain peptides failed to sensitize *PUMA*-deficient cells to UV-induced apoptosis, which highlights the specificity of the PUMA BH3 domain in this pathway (fig. S4E). BAD or PUMA BH3 domain peptides enhanced staurosporine-induced apoptosis independently of p53, suggesting that PUMA may also coordinate with other stimuli to promote death (fig. S4D). Thus, although some BH3-only proteins directly activate proapoptotic multidomain Bcl-2 proteins, such as BAX, to induce MOMP, several BH3-only proteins, like PUMA, regulate specific apoptotic pathways by interfering with antiapoptotic Bcl-2 proteins such as Bcl-xL (*19, 20*).

Structural analyses indicate that p53 interacts with amino acid residues on Bcl-xL distinct from those that mediate the binding of Bcl-xL to BAK, BAD, or BIM (*21–24*). Consistent with this, a mutant of Bcl-xL, Bcl-xL[G138A], which does not bind to proapoptotic multidomain Bcl-2 proteins or BH3-only proteins (*25*), and Bcl-xL[wt] associated equally with p53, whereas BAX and PUMA bound Bcl-xL[wt] but not Bcl-xL[G138A] (Fig. 3A). Likewise, PUMA protein and PUMA BH3 domain peptide failed to disrupt the Bcl-xL[G138A]·p53 complex in vitro (Fig. 3B), whereas both effectively released p53 from Bcl-xL[wt]. Tran-
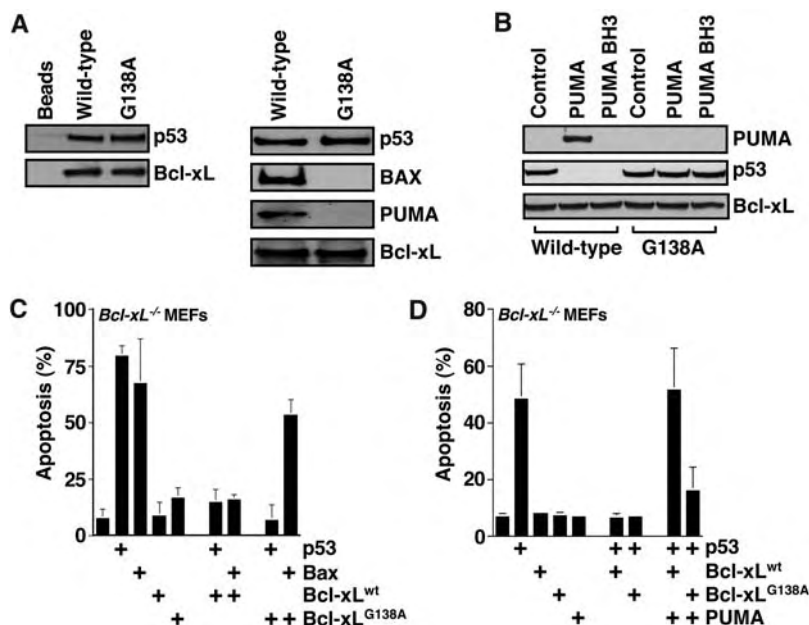

Fig. 3. Association of a BH3-domain binding mutant of Bcl-xL with p53 and inhibition of p53-induced death. (A) [6XHis]Bcl-xLΔC[wt] or [6XHis]Bcl-xLΔC[G138A] (10 ng) was combined with 10 ng of p53[wt] in phosphate-buffered saline (PBS) and incubated at 25°C for 2 hours. Ni[2+] beads were used to capture the complexed proteins before SDS-PAGE and Western blot analysis for p53 and Bcl-xL (left). "Beads" represents Ni[2+] beads incubated alone with p53[wt] in PBS. (Right) Full-length BAX and PUMA (10 ng) were also examined. (B) Recombinant 10 nM [6XHis]Bcl-xL[wt or G138A]·p53 complexes were added to 50 nM PUMA or 10 μM PUMA BH3 peptide and analyzed by Western blot for association with p53 or PUMA. (C and D) E1A/H-rasG12V–expressing *Bcl-xL[−/−]* MEFs were transiently transfected with the indicated plasmids (pcDNA3.control, pcDNA3.p53[wt], pcDNA3.BAX, pCMV.FLAG-PUMA, pcDNA3.Bcl-xL[wt] or pcDNA3.Bcl-xL[G138A]), and apoptosis was analyzed 24 hours later. Error bars represent ±SD.

Fig. 4. Bcl-xL[G138A] reconstitutes Bcl-xL deficiency for p53-dependent apoptosis. (A) E1A/H-rasG12V–transformed wild-type MEFs expressing pBABE.control, or *Bcl-xL[−/−]* MEFs stably expressing either pBABE.control, pBABE.Bcl-xL[wt], or pBABE.Bcl-xL[G138A] were treated with the indicated doses of cycloheximide (top) or UV (bottom) and analyzed 24 hours later for apoptosis. Error bars represent ±SD. (B) Proteins from the cytosol of 5 mJ/cm² UV-treated E1A/H-rasG12V–transformed *Bcl-xL[−/−]* MEFs stably expressing either pBABE.Bcl-xL[wt] or pBABE.Bcl-xL[G138A] were immunoprecipitated for Bcl-xL. Bcl-xL–coprecipitated p53 (Bound) and PUMA, and non-Bcl-xL–associated p53 (Free) and PUMA in the cytosol after Bcl-xL immunoprecipitation were analyzed by SDS-PAGE and Western blot. (C) Proposed model of p53-dependent, DNA damage–induced apoptosis. After p53 stabilization, p53 accumulates in the nucleus to directly regulate the expression of proapoptotic genes, such as *BAX* and *PUMA*. Likewise, p53 accumulates in the cytoplasm and directly binds to Bcl-xL. Once p53-dependent expression of *PUMA* occurs, PUMA binds to Bcl-xL, which releases p53 to directly activate BAX and induce MOMP.
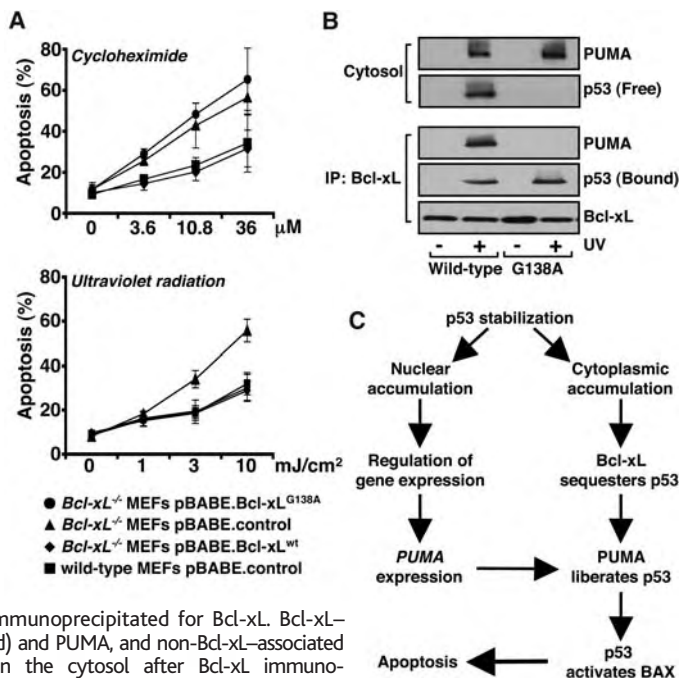
sient expression of either Bcl-xL^{wt} or Bcl-xL^{G138A} specifically blocked apoptosis induced by exogenous expression of p53 in E1A- and H-rasG12V–expressing *Bcl-xL$^{-/-}$* MEFs. In contrast, Bcl-xL^{G138A} failed to inhibit apoptosis induced by BAX expression (Fig. 3C). The ability of Bcl-xL^{wt} to block p53-induced apoptosis, but not that of Bcl-xL^{G138A}, was overcome by transient expression of PUMA, leading to apoptosis (Fig. 3D, fig. S3C). Thus, the association between p53 and Bcl-xL appears not to be the signal to induce death (that is, p53 does not inactivate Bcl-xL to induce apoptosis). Rather, this complex requires an additional proapoptotic stimulus (such as PUMA) to release p53 and engage cell death.

The effect of Bcl-xL^{G138A} on p53-dependent apoptosis was further examined in E1A- and H-rasG12V–expressing *Bcl-xL$^{-/-}$* MEFs reconstituted to stably express Bcl-xL^{wt} or Bcl-xL^{G138A} at normal levels (Fig. 4A, fig. S3D) (*26*). *Bcl-xL$^{-/-}$* MEFs expressing Bcl-xL^{G138A} demonstrated no protection against p53-independent death stimulation induced by cycloheximide, and their response was similar to that of *Bcl-xL$^{-/-}$* control cells. Stable expression of Bcl-xL^{wt} inhibited cycloheximide-induced apoptosis, and Bcl-xL^{G138A} expression also conferred resistance to p53-dependent, UV-induced apoptosis, suggesting that Bcl-xL^{G138A} can efficiently and specifically inhibit p53-dependent death (Fig. 4A). In extracts from the stable cell lines, immunoprecipitation of Bcl-xL^{wt} or Bcl-xL^{G138A} captured cytosolic p53 (Fig. 4B). Cytosolic p53 that not was immunoprecipitated with Bcl-xL was observed only in UV-treated Bcl-xL^{wt} cells (Fig. 4B). In

both cases, PUMA was expressed after UV treatment, but coprecipitated only with Bcl-xL^{wt}, not Bcl-xL^{G138A}. Because Bcl-xL^{G138A} binds cytoplasmic p53, but not BAX or PUMA, and is similarly deficient in binding other proapoptotic BH3-only proteins, the simplest interpretation of these results is that Bcl-xL^{G138A} inhibits p53-dependent apoptosis by sequestering cytoplasmic p53 and prevents it from activating BAX to promote MOMP. The nuclear functions of p53 are unaffected by Bcl-xL^{wt} or Bcl-xL^{G138A}, as evidenced by the equivalent expression of PUMA after UV treatment (Fig. 4B). Thus, cytoplasmic p53 appears to mediate apoptosis under physiological conditions.

Our studies support a model of p53-dependent, DNA damage–induced apoptosis that includes both nuclear and cytoplasmic functions of p53 (Fig. 4C). After p53 stabilization, p53 accumulates in the nucleus to directly regulate the expression of proapoptotic genes, such as *BAX* and *PUMA*. p53 also accumulates in the cytoplasm, directly binds to Bcl-xL, and awaits a secondary death signal. Once p53-dependent expression of *PUMA* occurs, PUMA binds to Bcl-xL, releasing p53 to directly activate BAX and induce MOMP. Our results may explain why *Bcl-xL*–deficient animals display sensitivity to DNA damage, whereas *Puma* deficiency promotes resistance to numerous p53-dependent apoptotic stimuli (*16, 26, 27*).

References and Notes

1. C. A. Schmitt *et al.*, *Cancer Cell* **1**, 289 (2002).
2. L. B. Owen-Schaub *et al.*, *Mol. Cell. Biol.* **15**, 3032 (1995).
3. M. Muller *et al.*, *J. Exp. Med.* **188**, 2033 (1998).
4. T. Miyashita, J. C. Reed, *Cell* **80**, 293 (1995).
5. E. Oda *et al.*, *Science* **288**, 1053 (2000).
6. K. Nakano, K. H. Vousden, *Mol. Cell* **7**, 683 (2001).
7. J. K. Sax *et al.*, *Nat. Cell Biol.* **4**, 842 (2002).
8. M. Mihara *et al.*, *Mol. Cell* **11**, 577 (2003).
9. J. E. Chipuk *et al.*, *Science* **303**, 1010 (2004).
10. P. Dumont, J. I. Leu, A. C. Della Pietra III, D. L. George, M. Murphy, *Nat. Genet.* **33**, 357 (2003).
11. N. N. Danial, S. J. Korsmeyer, *Cell* **116**, 205 (2004).
12. X. Chen, L. J. Ko, L. Jayaraman, C. Prives, *Genes Dev.* **10**, 2438 (1996).
13. Y. Haupt, S. Rowan, E. Shaulian, K. H. Vousden, M. Oren, *Genes Dev.* **9**, 2170 (1995).
14. C. Caelles, A. Helmberg, M. Karin, *Nature* **370**, 220 (1994).
15. J. E. Chipuk, U. Maurer, D. R. Green, M. Schuler, *Cancer Cell* **4**, 371 (2003).
16. A. Villunger *et al.*, *Science* **302**, 1036 (2003).
17. J. R. Jeffers *et al.*, *Cancer Cell* **4**, 321 (2003).
18. J. Yu, Z. Wang, K. W. Kinzler, B. Vogelstein, L. Zhang, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1931 (2003).
19. T. Kuwana *et al.*, *Mol. Cell* **17**, 525 (2005).
20. L. Chen *et al.*, *Mol. Cell* **17**, 393 (2005).
21. M. Sattler *et al.*, *Science* **275**, 983 (1997).
22. A. M. Petros *et al.*, *Protein Sci.* **9**, 2528 (2000).
23. X. Liu, S. Dai, Y. Zhu, P. Marrack, J. W. Kappler, *Immunity* **19**, 341 (2003).
24. A. M. Petros, A. Gunasekera, N. Xu, E. T. Olejniczak, S. W. Fesik, *FEBS Lett.* **559**, 171 (2004).
25. E. Yang *et al.*, *Cell* **80**, 285 (1995).
26. N. Motoyama *et al.*, *Science* **267**, 1506 (1995).
27. B. J. Klocke, C. B. Latham, C. D'Sa, K. A. Roth, *Cell Death Differ.* **9**, 1063 (2002).
28. We thank P. Juin, J. Claude Martinou, K. Roth, A. Strasser, B. Sullivan, A. Villunger, B. Vogelstein, K. Vousden, and G. Wahl for their generosity with reagents and expertise. This work was supported by grants from the Leukemia and Lymphoma Society (D.R.G.) and the National Institutes of Heath (D.R.G. and D.D.N.). J.E.C. is supported by an Individual National Research Service Award.

Lymphocyte Sequestration Through S1P Lyase Inhibition and Disruption of S1P Gradients

Susan R. Schwab,[1] João P. Pereira,[1] Mehrdad Matloubian,[1] Ying Xu,[1] Yong Huang,[2] Jason G. Cyster[1]*

Lymphocyte egress from the thymus and from peripheral lymphoid organs depends on sphingosine 1-phosphate (S1P) receptor-1 and is thought to occur in response to circulatory S1P. However, the existence of an S1P gradient between lymphoid organs and blood or lymph has not been established. To further define egress requirements, we addressed why treatment with the food colorant 2-acetyl-4-tetrahydroxybutylimidazole (THI) induces lymphopenia. We found that S1P abundance in lymphoid tissues of mice is normally low but increases more than 100-fold after THI treatment and that this treatment inhibits the S1P-degrading enzyme S1P lyase. We conclude that lymphocyte egress is mediated by S1P gradients that are established by S1P lyase activity and that the lyase may represent a novel immunosuppressant drug target.

Lymphocyte egress from lymphoid organs is critical for immune surveillance and immune effector function and depends on intrinsic expression of sphingosine 1-phosphate receptor-1 (S1P$_1$), a seven-transmembrane receptor that binds S1P (*1–4*). S1P is abundant in blood, and it has been suggested that lymphocytes exit from lymphoid tissues into blood or lymph in

response to S1P gradients (*5, 6*). However, all cell types generate S1P intracellularly during sphingolipid metabolism, making it unclear which factors regulate the S1P levels in tissues versus in plasma and lymph (*7, 8*). To further define requirements for lymphocyte exit, we followed up on a finding made during toxicity testing of caramel food colorants. In these tests, a component of caramel color III, 2-acetyl-4-tetrahydroxybutylimidazole (THI), induced lymphopenia (*9–11*) and caused mature T cells to accumulate in the thymus, apparently as a result of reduced egress (*12, 13*).

We observed that in addition to causing an accumulation of mature T cells in the thymus, THI inhibited lymphocyte egress from lymph nodes (Fig. 1). Mice treated for three days with THI (*14*) were depleted of CD4 and CD8 T cells and, to a lesser degree, B cells in both

[1]Howard Hughes Medical Institute and Department of Microbiology and Immunology and [2]Drug Studies Unit, Department of Biopharmaceutical Sciences, University of California San Francisco, 513 Parnassus Avenue, San Francisco, CA 94143–0414, USA

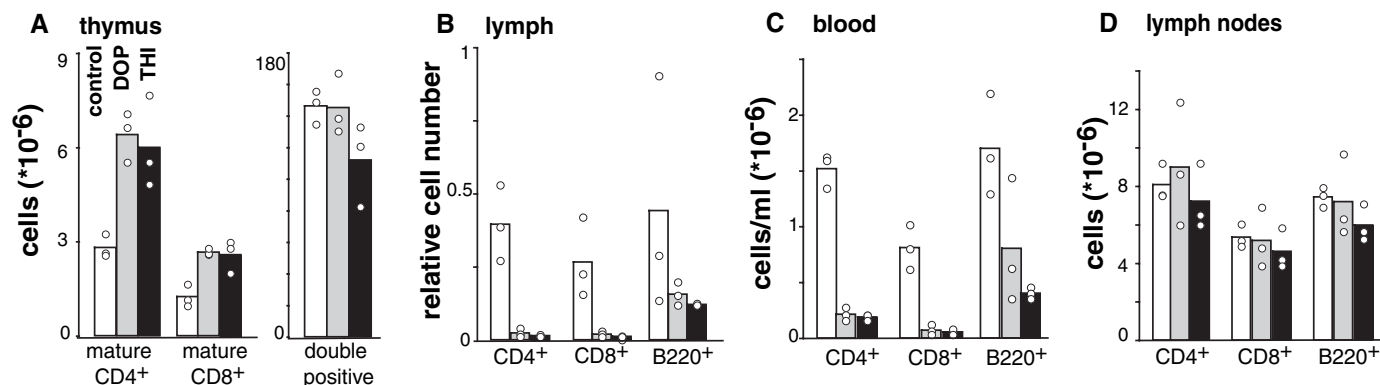*To whom correspondence should be addressed. E-mail: cyster@itsa.ucsf.edu

Fig. 1. THI and DOP cause accumulation of mature cells in the thymus and loss of lymphocytes from the lymph and blood. (**A** to **D**) Lymphocyte numbers in the indicated tissues from mice given drinking water alone (control, white bars), given water containing 30 mg of DOP per liter (gray bars), or given water containing 50 mg of THI per liter (black bars) for three days. Mature thymocytes (A) were defined as CD4$^+$ or CD8$^+$ and CD69^{lo}CD62L^{hi}, and double-positive thymocytes were defined as CD4$^+$ and CD8$^+$. Points indicate values from individual mice and bars indicate the average. When data from three experiments were combined, the increase in mature thymocytes and decrease in blood and lymph lymphocytes were statistically significant ($P < 0.05$), with the exception of the decrease in B220$^+$ cells in the lymph of DOP-treated mice.
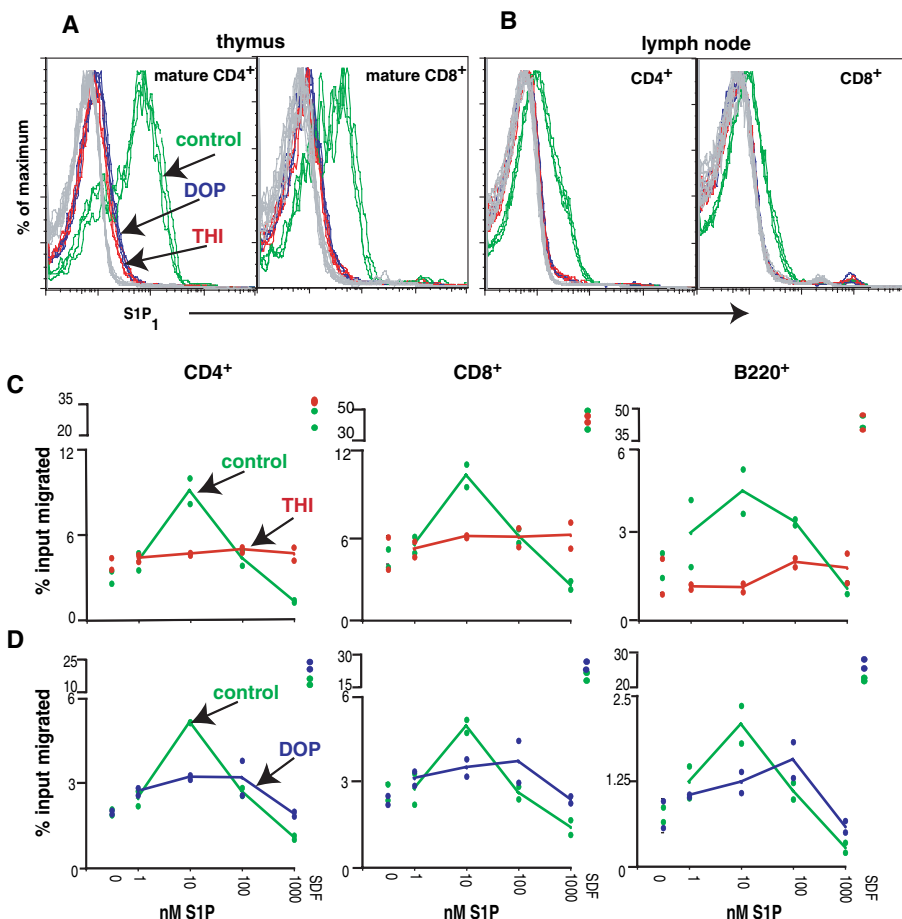

Fig. 2. THI and DOP treatment lead to reduced S1P$_1$ expression and impaired ability to migrate to S1P. (**A**) Mature single-positive thymocytes and (**B**) peripheral lymph node cells were stained for S1P$_1$ expression. Green lines show staining of cells from three control-treated mice, blue from three DOP-treated mice, and red from three THI-treated mice. Gray lines show staining of all nine cell types with a control antibody. Staining is representative of at least three experiments. (**C** and **D**) Splenocytes from THI- (red), DOP- (blue), or control- (green) treated mice were tested for their ability to migrate to the indicated concentrations of S1P or to 0.3 μg of stromal-cell derived factor (CXCL12) per ml of medium. Individual points show migration in replicate wells. When data from three experiments were compiled, the impaired migration of CD4$^+$ and CD8$^+$ T cells from DOP- and THI-treated mice to 10 nM S1P was statistically significant ($P < 0.05$). The impaired migration of B220$^+$ cells was significant for THI ($P = 0.05$) but not for DOP.

lymph and blood, without losing cells in lymph nodes (Fig. 1, B to D). Because naïve lymphocytes enter lymph only by exiting lymph nodes and Peyer's patches, this result indicated that THI blocked exit. During toxicology studies on THI, an investigation of inconsistencies among data from different laboratories revealed that the sequestration was prevented by excess dietary vitamin B$_6$ (*11*). We therefore also examined the effect of treating mice with 4′ deoxypyridoxine (DOP), a vitamin-B$_6$ antagonist (*15*), and found that DOP mimicked THI in causing a build up of mature cells in the thymus, depleting T and B lymphocytes from lymph, and inducing lymphopenia (Fig. 1). Adoptively transferred cells disappeared from the blood and lymph during THI or DOP treatment but returned after treatment was stopped, indicating that the compounds were not causing a loss of cells due to toxicity (fig. S1) (*16*).

Because S1P$_1$ is essential for lymphocyte egress from the thymus and secondary lymphoid organs, we tested whether THI and DOP were acting by altering S1P$_1$ function. Flow cytometric analysis revealed that S1P$_1$ expression was almost undetectable on mature CD4 or CD8 single-positive thymocytes (CD4$^+$ or CD8$^+$) from THI- and DOP-treated mice but was readily detected on control thymocytes (Fig. 2A). Similarly, THI and DOP treatment caused a loss of surface S1P$_1$ on lymph node CD4 and CD8 T cells (Fig. 2B). S1P$_1$ mediates chemotaxis of lymphocytes to S1P. Consistent with the loss of surface expression, neither CD4 nor CD8 splenic T cells from THI- and DOP-treated mice could migrate in response to S1P, and a similar, although less severe, impairment was seen in follicular B cells (Fig. 2, C and D).

Cells internalize S1P$_1$ upon exposure to S1P (*17*), and incubation of freshly isolated thymocytes with as little as 0.5 nM S1P down-modulated surface S1P$_1$ expression (Fig. 3A). We therefore considered the possibility that
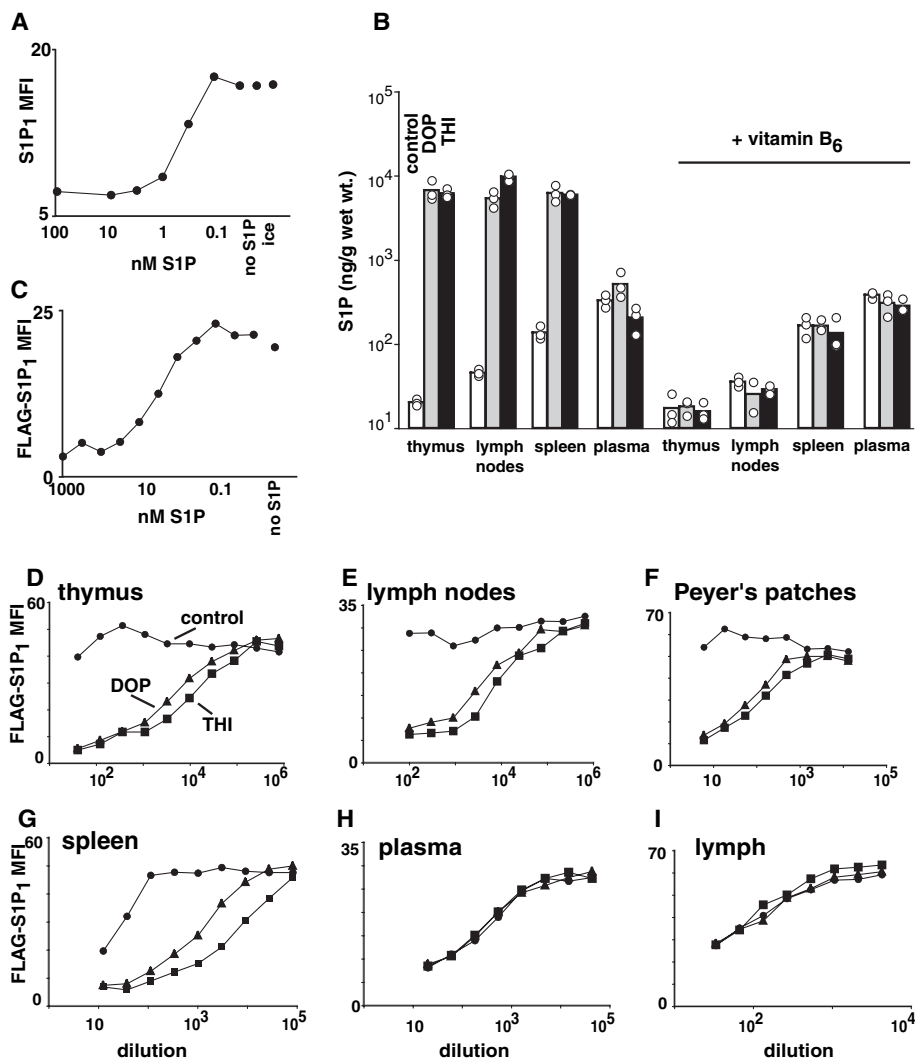
Fig. 3. THI and DOP increase lymphoid tissue S1P concentrations. (**A**) Mature CD4$^+$ thymocytes from an untreated mouse were incubated in the indicated concentration of S1P, and S1P$_1$ mean fluorescence intensity (MFI) levels were measured. (**B**) S1P levels in the indicated tissues from control- (open bars), DOP- (gray), or THI- (black) treated mice, measured by LC/MS/MS. Points represent values from individual mice, and bars represent the average. In some cases, the mice also received vitamin B$_6$ (pyridoxine HCl, 450 mg per liter of water), as indicated. The thymic, lymph node, and spleen S1P concentrations in THI- and DOP-treated mice were significantly different from the control ($P < 0.05$). (**C**) FLAG-S1P$_1$-expressing WEHI231 cells were incubated in the indicated concentration of S1P, and FLAG-S1P$_1$ MFI levels were measured. (**D** to **G**) Extracts enriched in interstitial fluid from the indicated tissues were prepared (14) and titrated onto FLAG-S1P$_1$ WEHI231 cells. The x axis shows dilution of total tissue from which the extract was made, assuming that the tissue density is 1 g/ml. (**H** and **I**) As in (C), but x axis shows dilution of plasma or lymph. Maximum FLAG-S1P$_1$ levels vary among experiments because they were performed on different days. All data are representative of at least three experiments, with the exception of MS analysis of vitamin-B$_6$ treated mice and S1P measurement in Peyer's patches.

THI and DOP were causing S1P concentrations within lymphoid tissues to increase. The extraction of lipids from plasma and whole tissue and the quantification of total S1P by LC/MS/MS (liquid chromatography followed by tandem mass spectrometry) (14) revealed that lipid levels in plasma were approximately 330 ng/g (860 nM) and were not changed by THI or DOP treatment (Fig. 3B and fig. S2). In contrast, tissue S1P concentrations rose markedly from ~20, 40, and 150 ng/g of tissue in the thymus, lymph nodes, and spleen, respectively, to 5,000 to 10,000 ng/g in each

of these tissues after treatment (Fig. 3B). Total tissue analysis measures intracellular and extracellular S1P. To selectively determine the receptor-accessible extracellular S1P concentration, we took two steps. First, we developed a bioassay in which S1P-induced down-modulation of FLAG-tagged S1P$_1$ expression in WEHI231 cells could be measured. When S1P was titrated onto the cells, detectable receptor down-regulation at 1 nM was observed (Fig. 3C). Second, we prepared tissue extracts enriched in interstitial fluid by disaggregating lymphoid organs to a single cell

suspension in isotonic medium and removing the cells by centrifugation. Because our goal was to gain an indication of the bioavailable S1P, we did not further purify the lipids, and it is possible that our measurements underestimate total extracellular S1P, because some may be sequestered. When we titrated these extracts onto the WEHI231 cells, we found that S1P was present in low to undetectable amounts in extracts from control thymus tissue, lymph nodes, and Peyer's patches (Fig. 3, D to F). The detection of S1P in the control spleen extract is consistent with this organ being partly composed of plasma-rich red pulp (Fig. 3G). S1P was abundant in plasma, and amounts in lymph were lower but within sixfold of those in plasma (Fig. 3, H and I). Analysis of extracts from THI- and DOP-treated mice revealed that S1P abundance was increased more than 1000-fold in the thymus, lymph node, and Peyer's patch extracts and more than 100-fold in spleen extracts (Fig. 3, D to G). In contrast, there was no difference in plasma and lymph S1P concentrations between the treated and untreated mice (Fig. 3, H and I).

Of the enzymes involved in determining S1P abundance, S1P lyase, a cytoplasmic enzyme that degrades S1P into phosphoethanolamine and 2-hexadecanal, represented an appealing candidate as a potential target for THI and DOP, because it is dependent on vitamin B$_6$ (8, 18, 19). Mice that had received THI or DOP combined with a 10-fold molar excess of vitamin B$_6$ in their drinking water showed no accumulation of mature thymocytes, induction of lymphopenia, down-modulation of S1P$_1$, or increase in lymphoid tissue S1P levels (Fig. 3B and fig. S3). We therefore tested whether S1P lyase activity in the thymus was affected by THI or DOP treatment. Thymic lysates from treated and untreated mice were incubated with tritiated dhS1P (dihydrosphingosine 1-phosphate), and the products were separated by thin-layer chromatography (14). Lyase activity, measured by the generation of hexadecanal, was severely inhibited in the thymus of DOP- and THI-treated mice (Fig. 4, A and B). Excess vitamin B$_6$ overcame the effect of THI or DOP and restored lyase activity (Fig. 4, A and B). These findings suggest that THI and DOP cause increased lymphoid tissue S1P abundance because of the inhibition of S1P lyase.

To further test whether a reduction in S1P lyase activity could account for the increased tissue S1P abundance and disruption in egress, we reduced hematopoietic cell lyase expression by RNA interference (RNAi)–mediated knockdown (14). Hematopoietic stem cells (HSCs) were transduced with a retrovirus containing a lyase RNAi-inducing small hairpin RNA (shRNA) and a GFP reporter (ZsGreen green fluorescent protein), or the reporter alone. The transduced HSCs were used to reconstitute irradiated lymphocyte-deficient (RAG2$^{-/-}$) mice. Quantitative polymerase chain reaction (PCR)

analysis revealed that S1P lyase mRNA abundance was reduced by over 80% in single-positive thymocytes with high amounts of GFP (GFP[hi]) from mice reconstituted with HSCs expressing the shRNA (Fig. 4C and fig. S4). Lyase expression was reduced ~threefold in a thymocyte suspension composed predominantly of hematopoetic cells, and by about 30%

in total thymus tissue (Fig. 4C). Threefold reductions in lyase expression were observed in the spleen (Fig. 4C). Lyase knockdown led to an accumulation of mature single-positive cells in the thymus, mirrored by a T cell deficiency in the spleen and blood (Fig. 4D). Surface S1P$_1$ was partially down-regulated on mature thymocytes, and S1P abundance in the

thymus and spleen was elevated as determined by both LC/MS/MS analysis and bioassay (Fig. 4, F to I). The generally less severe phenotype that exists in shRNA-expressing mice, as opposed to that in THI- or DOP-treated mice, most likely occurs because THI and DOP reduce S1P lyase activity in both hematopoietic and nonhematopoietic cells,
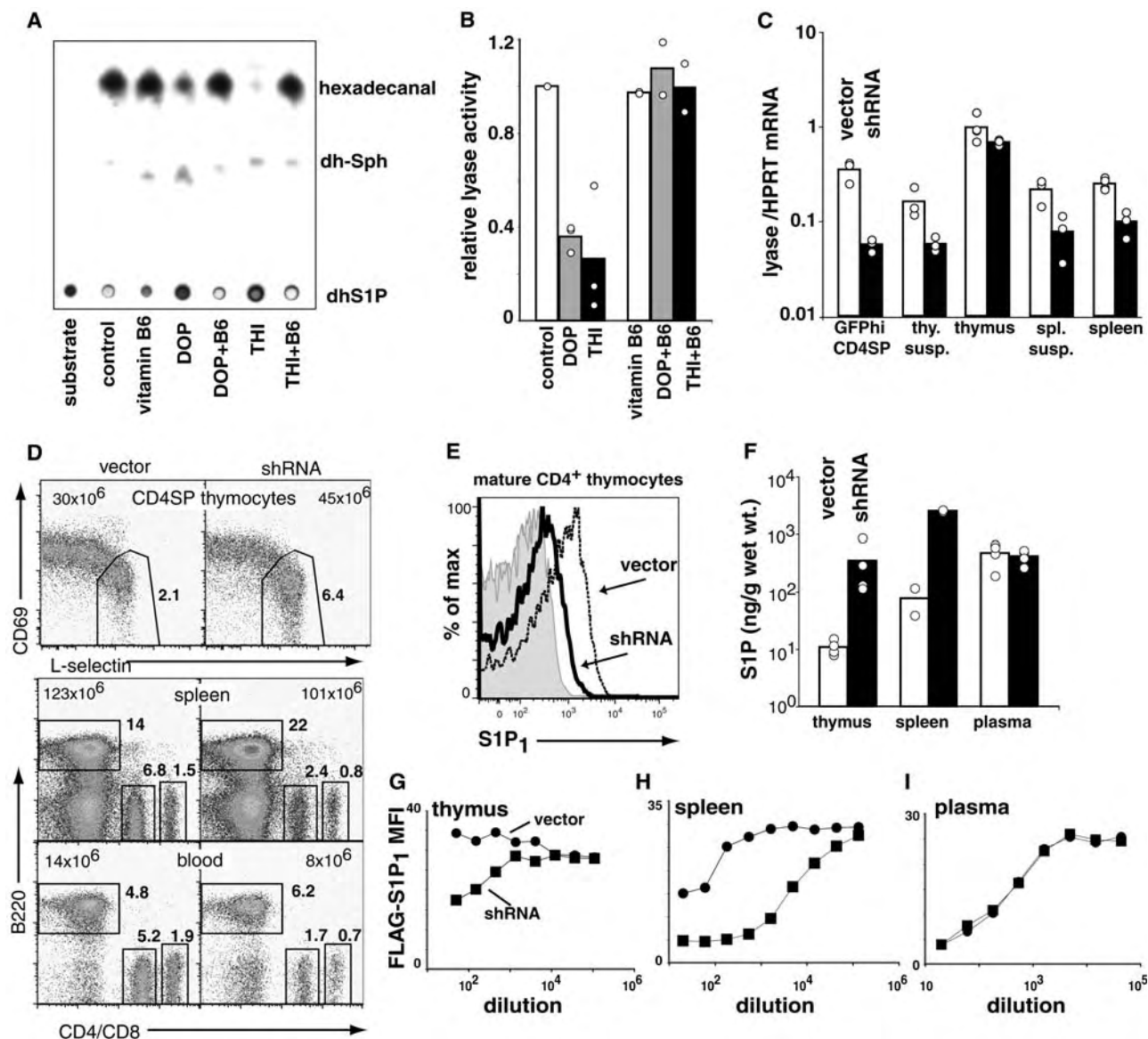

Fig. 4. THI and DOP inhibit S1P lyase, and RNAi-mediated S1P lyase knockdown recapitulates the THI- and DOP-induced phenotype. (**A** and **B**) Thymic lysates from mice treated as indicated (B6, vitamin B$_6$) were incubated with tritiated dhS1P. The products were separated by thin-layer chromatography and visualized by autoradiography (A). The lyase product hexadecanal and its metabolites hexadecanol and palmitic acid have a similar R_f (where R_f is the distance moved by the lipid divided by the distance from the origin to the solvent front) (*19*) and were not distinguished in our assay. dh-Sph, dihydrosphingosine. In (B), quantitative analysis by PhosphorImager (GE Healthcare, Chalfont St. Giles, UK) of thymic lyase activity in three experiments is plotted as percent of control activity. The activity reduction in lysates from THI- and DOP-treated mice was statistically significant ($P <$ 0.05). (**C**) Quantitative real-time PCR analysis of lyase relative to hypoxanthine-guanine phosphoribosyltransferase (HPRT) transcript abundance in the indicated tissues from mice reconstituted with HSC transduced with a retrovirus encoding the shRNA and a GFP reporter (black bars), or the

GFP reporter alone (open bars). GFP[hi] CD4SP indicates GFP-high, CD4$^+$ CD8$^-$ thymocytes; thy. susp., thymic suspension; thymus, total thymus; spl. susp., splenic suspension; spleen, total spleen. Bars show average, points show values from individual animals. The RNA levels in GFP[hi] thymocytes, thymic suspension, splenic suspension, and total spleen of mice receiving shRNA were significantly different from the control ($P <$ 0.05). (**D**) Flow cytometric analysis of representative thymuses, spleens, and blood, gated by size on lymphocytes. Numbers in the upper corners show the total number of cells in the tissue or cells per milliliter in blood, and the numbers next to gates show the percent of total cells in the gate. (**E**) Representative flow cytometric analysis of S1P$_1$ expression on mature CD4$^+$ thymocytes. Shaded histograms show staining of both cell types with a control antibody. (**F**) S1P levels in total thymus, spleen, and plasma measured by LC/MS/MS. The thymic and spleen S1P concentrations in mice that received shRNA were significantly different from the control ($P <$ 0.05). (**G** to **I**) Representative bioassay measurement of S1P levels in the indicated tissue extracts, as in Fig. 3.

whereas the knockdown approach affects only a fraction of the hematopoietic cells. Nonetheless, selective RNAi-mediated knockdown of S1P lyase recapitulated key aspects of the THI- and DOP-induced phenotype.

Although there are several mechanisms by which S1P levels can be regulated (8), we provide evidence that within lymphoid tissues, S1P lyase expressed by hematopoetic cells has an essential role in degrading S1P and keeping tonic S1P concentrations low. THI, DOP, and shRNA expression may each have effects unrelated to S1P lyase, but the finding that three independent means of inhibiting the lyase all elevate lymphoid S1P concentrations firmly establishes a role for the lyase in controlling S1P levels. These findings provide evidence that there is normally a large differential in bioavailable S1P abundance between the circulatory fluids—plasma and lymph—and the lymphoid organs. The increased S1P concentration induced by lyase inhibition provides an explanation for the loss of surface $S1P_1$ from mature lymphocytes, and reduced $S1P_1$ function could account for the inhibition of egress from the thymus and peripheral lymphoid organs. It also is possible that exit could be inhibited by a loss of guidance information provided by the S1P gradient or by S1P-mediated changes in other cell types, such as endothelial cells (20). A loss of guidance information has been suggested as an ex-

planation for the germ-cell migration defect in *wunen* mutant *Drosophila* embryos, which lack a lipid phosphatase (21). Chronic vitamin-B_6 deficiency is associated with lymphopenia and immune deficiency (11, 22), effects that might in part be explained by reduced lyase activity and defective lymphocyte egress. Although our findings demonstrate that THI-ingestion leads to an inhibition of S1P lyase activity, further studies will be needed to define whether THI can bind the lyase directly or whether it first becomes modified in vivo or acts by an indirect mechanism. The disruption of $S1P_1$ function by small molecules such as FTY720 is currently being tested as a mode of immunosuppression in humans (6, 20). Our studies suggest that S1P lyase represents a novel target for immunosuppressive drugs.

References and Notes

1. S. Mandala *et al.*, *Science* **296**, 346 (2002).
2. V. Brinkmann *et al.*, *J. Biol. Chem.* **277**, 21453 (2002).
3. M. Matloubian *et al.*, *Nature* **427**, 355 (2004).
4. M. L. Allende, J. L. Dreier, S. Mandala, R. L. Proia, *J. Biol. Chem.* **279**, 15396 (2004).
5. J. G. Cyster, *Annu. Rev. Immunol.* **23**, 127 (2005).
6. H. Rosen, E. J. Goetzl, *Nat. Rev. Immunol.* **5**, 560 (2005).
7. Y. Yatomi, Y. Ozaki, T. Ohmori, Y. Igarashi, *Prostaglandins Other Lipid Mediat.* **64**, 107 (2001).
8. H. Le Stunff, S. Milstien, S. Spiegel, *J. Cell. Biochem.* **92**, 882 (2004).
9. J. G. Evans, K. R. Butterworth, I. F. Gaunt, P. Grasso, *Food Cosmet. Toxicol.* **15**, 523 (1977).
10. I. F. Gaunt, A. G. Lloyd, P. Grasso, S. D. Gangolli, K. R. Butterworth, *Food Cosmet. Toxicol.* **15**, 509 (1977).
11. E. J. Sinkeldam, A. P. de Groot, H. van den Berg, C. I. Chappel, *Food Chem. Toxicol.* **26**, 195 (1988).
12. S. J. Gobin, A. J. Paine, *Thymus* **20**, 17 (1992).
13. R. Gugasyan, A. Coward, L. O'Connor, K. Shortman, R. Scollay, *Immunology* **93**, 398 (1998).
14. Materials and methods are available as supporting material on *Science* Online.
15. S. P. Coburn, *The Chemistry and Metabolism of 4′-deoxypyridoxine* (CRC Press, Boca Raton, FL, 1981).
16. M. G. Bradbury, K. V. Doherty, C. R. Parish, A. B. Lyons, *Immunology* **87**, 80 (1996).
17. M. J. Lee *et al.*, *Science* **279**, 1552 (1998).
18. J. D. Saba, T. Hla, *Circ. Res.* **94**, 724 (2004).
19. P. P. Van Veldhoven, *Methods Enzymol.* **311**, 244 (2000).
20. V. Brinkmann, J. G. Cyster, T. Hla, *Am. J. Transplant.* **4**, 1019 (2004).
21. A. D. Renault, Y. J. Sigal, A. J. Morris, R. Lehmann, *Science* **305**, 1963 (2004).
22. A. Trakatellis, A. Dimitriadou, M. Trakatelli, *Postgrad. Med. J.* **73**, 617 (1997).
23. We thank R. Pappu, other members of S. Coughlin's laboratory, and H. E. Witkowska for advice on S1P measurement; J. Saba, M. Hoyt, and P. Coffino for enzymology discussions; M. Jain and J. Cox for help with thin-layer chromatography; C. Lo and T. Okada for technical advice; and C. Allen, N. Haynes, and L. Shiow for manuscript comments. S.R.S. is a Damon Runyon Cancer Research Foundation fellow. J.G.C. is a Howard Hughes Medical Institute investigator. This work was supported in part by grants AI40098 and AI45073 from NIH.

Plague Bacteria Target Immune Cells During Infection

Melanie M. Marketon,[1]* **R. William DePaolo,**[2]* **Kristin L. DeBord,**[1] **Bana Jabri,**[2] **Olaf Schneewind**[1]†

The plague is caused by the bacterium *Yersinia pestis*. Plague bacteria are thought to inject effector Yop proteins into host cells via the type III pathway. The identity of the host cells targeted for injection during plague infection is unknown. We found, using Yop β-lactamase hybrids and fluorescent staining of live cells from plague-infected animals, that *Y. pestis* selected immune cells for injection. In vivo, dendritic cells, macrophages, and neutrophils were injected most frequently, whereas B and T lymphocytes were rarely selected. Thus, it appears that *Y. pestis* disables these cell populations to annihilate host immune responses during plague.

Yersinia pestis, the causative agent of plague or black death, harbors a virulence plasmid that encodes a type III secretion machine and its Yop protein substrates (1, 2). The essential contribution of type III secretion to the pathogenesis of plague was revealed by comparing lethal infectious doses of wild-type and mutant strains (3). *Yersinia* type III injection of Yop proteins into tissue culture cells has been detected with fluorescent microscopy, adenylate cyclase or Elk tag fusions, and fractionation techniques (4–7). These technologies, however, have not been useful for measuring the selection of host cells as targets of type III injection during infection.

CCF2-AM, a β-lactamase substrate, has been used to detect bacterial type III reporter injection into tissue culture cells (8). CCF2-AM is a membrane-permeant ester with two fluorophores attached to cephalosporin that exhibit fluorescence resonance energy transfer (FRET). Excitation of coumarin (409 nm) results in green

fluorescence emission from fluorescein (520 nm) in intact CCF2-AM (9). β-Lactamase cleaves CCF2-AM, thereby disrupting FRET and establishing blue fluorescence emission. To investigate the usefulness of β-lactamase as a reporter for in vivo target cell selection by *Y. pestis*, we transformed plasmids carrying translational fusions between YopM or glutathione *S*-transferase (GST) and the mature domain of TEM-1 β-lactamase (YopM-Bla and GST-Bla, respectively) into *Y. pestis* strain KIM D27 (10). Bacterial cultures were induced for type III secretion via the depletion of calcium at 37°C and then centrifuged to separate the extracellular medium from bacterial cells (11). Immunoblotting of cell-associated and secreted proteins identified YopM-Bla in the extracellular medium of KIM D27 cultures (Fig. 1A). In contrast, GST-Bla or RpoA, the α subunit of RNA polymerase, were not secreted. Disruption of *yscU*, which encodes a secretion machine component, abrogates all type III secretion (12). Reporter plasmids transformed into *Y. pestis* strain CHI30, carrying a stable insertion of *MuAphP1* in *yscU*, revealed YopM-Bla only in bacterial cells (Fig. 1A) (13). Infection of tissue cultures with *Y. pestis* KIM D27 did not affect CCF2-AM–mediated green fluorescence, even though the HeLa cells rounded after 2 to 3 hours of infection as a result of the injection of Yop effectors (Fig. 1B). After infection of tissue culture cells with *Y. pestis* KIM

[1]Department of Microbiology, [2]Department of Pathology, University of Chicago, Chicago, IL 60637, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: oschnee@bsd.uchicago.edu
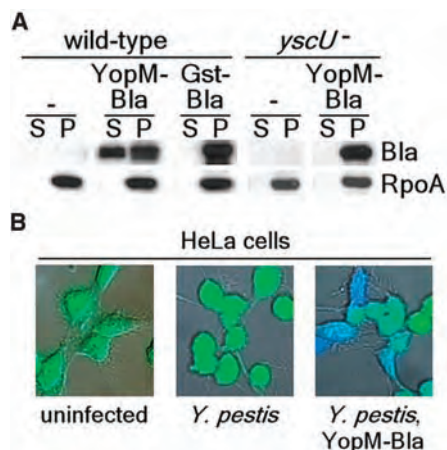
Fig. 1. (**A**) *Y. pestis* secretes YopM-Bla, but not GST-Bla, into the medium in a type III machinery (*yscU*)–dependent manner during in vitro secretion assays. Cultures were separated into supernatant (S) and pellet (P) fractions, and proteins were visualized by SDS–polyacrylamide gel electrophoresis (PAGE) and immunoblotting with antisera to β-lactamase (Bla) and RNA polymerase A (RpoA). (**B**) HeLa cells were infected with *Y. pestis* carrying either empty vector (center) or YopM-Bla (right), followed by staining with CCF2-AM and visualization of live cells by fluorescence microscopy. Blue and green fluorescence and differential interference contrast images were overlaid to obtain composite images.

D27 (pMM83, expressing YopM-Bla), cells fluoresced blue, revealing cleavage of the CCF2-AM cephalosporin moiety by the hybrid β-lactamase (Fig. 1B). GST-Bla was not injected into HeLa cells, as tissue cultures infected with KIM D27 (pMM91, expressing GST-Bla) generated only green fluorescence. YopM-Bla was not injected and blue fluorescence was not observed for HeLa cells infected with *yscU*, *yopD*, or *lcrV* mutants, because each of the disrupted genes is required for type III injection (fig. S1) (*14–16*). Thus, TEM-1 β-lactamase serves as a reporter for transport of hybrids by the *Y. pestis* type III secretion system.

To measure type III injection of primary cells, we isolated macrophages from the peritoneal cavity of mice, infected them with *Y. pestis*, and stained them with CCF2-AM. Infections with KIM D27 strains carrying pMM83 (YopM-Bla), but not pMM91 (GST-Bla), caused blue fluorescence of CCF2-AM–stained cells (Fig. 2). As observed with HeLa cells, blue fluorescence required functional *yscU* (Fig. 2) as well as *lcrV* and *yopD*. Thus, YopM-Bla and CCF2-AM staining can be used to measure selection of murine target cells by the *Yersinia* type III secretion system.

Y. pestis replicates within lymphoid tissues of experimental animals—that is, in the lymph nodes, bone marrow, liver, and spleen (*17, 18*). These organs are quickly filled with plague bacilli and are depleted of immune cells; they then undergo tissue necrosis, and the death of the infected animal follows (*19*). To test whether
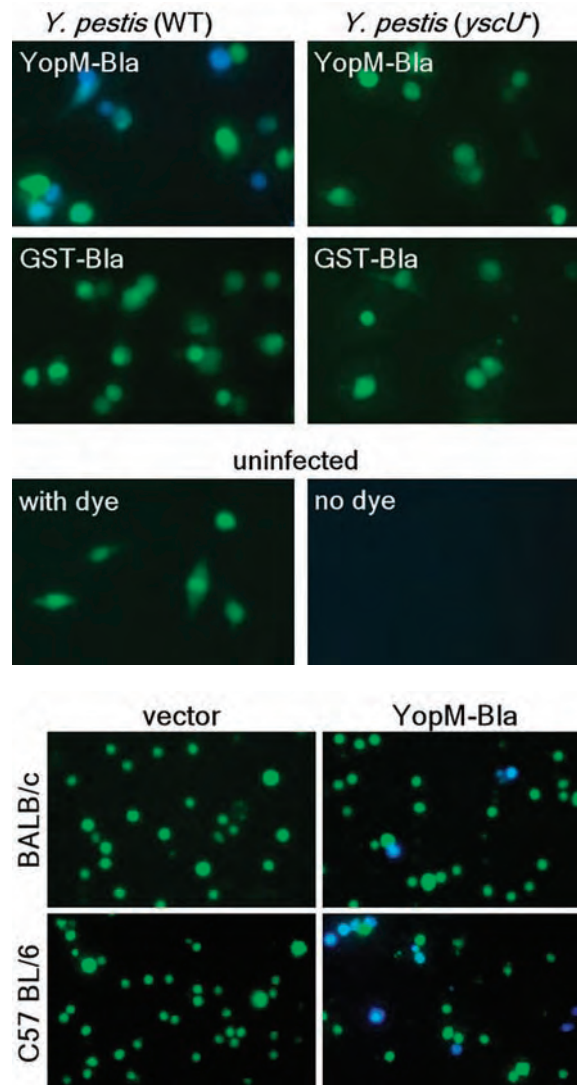

Fig. 2. Type III injection of peritoneal macrophages. Peritoneal macrophages were harvested from C57BL/6 mice and infected with *Y. pestis* strains (wild type or *yscU⁻*) carrying either YopM-Bla or GST-Bla. Live cells were stained with CCF2-AM and visualized by fluorescence microscopy.

Fig. 3. Mice were infected with 10⁵ CFU (BALB/c) or 10³ CFU (C57BL/6) of *Y. pestis* carrying either pHSG576 (vector) or pMM83 (YopM-Bla). After 2 days, infected spleens were removed, homogenized, and incubated with CCF2-AM; live cells were observed by fluorescence microscopy.

type III injection can be detected during plague infection, we infected mice intravenously with *Y. pestis* KIM D27 (pMM83) or *Y. pestis* KIM D27 (pHSG576, the vector control). All infections caused lethal plague disease within 4 days. At 2 days after infection, moribund mice were killed and spleens were removed. The tissue was homogenized, and cells were stained with CCF2-AM and viewed by fluorescence microscopy. Blue and green cells were present in spleens of BALB/c or C75BL/6 mice infected with KIM D27 (pMM83), whereas splenocytes from mice infected with KIM D27 (pHSG576) stained homogeneously green (Fig. 3). Both plasmids were stably maintained during infection and did not affect virulence (*13*).

We sought to determine the identity of cells targeted by the *Yersinia* type III pathway. Spleens of naïve C57BL/6 mice were homogenized and infected with *Y. pestis* KIM D27 carrying pMM83 (YopM-Bla), pMM85 (YopE-Bla), or pHSG576. After 3 hours, cells were stained with CCF2-AM and viewed by fluorescence microscopy. Some cells (e.g., erythrocytes) were not stained with CCF2-

AM and displayed no fluorescence. However, most splenocytes, either uninfected or infected with KIM D27 (pHSG576), fluoresced green. Splenocytes infected with plague strains KIM D27 (pMM83) and KIM D27 (pMM85) generated uniform blue fluorescence, signifying that these splenocytes had been injected with YopM-Bla or YopE-Bla (fig. S2). *Y. pestis* type III injection of splenocytes was not affected by treatment with cytochalasin D, which indicates that extracellular, but not phagocytosed, microbes generate blue fluorescent signals (fig. S3). To determine whether *Y. pestis* selected specific immune cells in vitro for type III injection, we isolated splenocytes and infected individual cell populations with *Y. pestis* strains KIM D27 carrying pMM83, pMM85, or pHSG576; some splenocytes were left uninfected. *Y. pestis* injected B cells, CD4⁺ T helper cells, CD8⁺ T killer cells, dendritic cells, or macrophages with YopM-Bla and YopE-Bla but not with GST-Bla (fig. S4).

To quantify injected cells, we analyzed CCF2-AM–stained cell populations by flow cytometry for the percentage of blue fluores-

Fig. 4. (A) Flow cytometric analysis of in vitro infections of purified cell populations with *Y. pestis* carrying pMM83 (YopM-Bla, white bars) or pMM85 (YopE-Bla, gray bars). The percentage of live blue cells for each cell type was plotted (data are means ± SEM of triplicate samples). (B) Mice were infected with 10^3 CFU of *Y. pestis* carrying either pHSG576 (vector) or pMM83 (YopM-Bla). Splenocytes were harvested and stained with CCF2-AM as well as cell-specific surface markers. The R1 gate was set by forward scatter (FS) versus side scatter (SS) and exclusion of propidium iodide–positive cells (top panel). R1-gated cells were analyzed for blue and green fluorescence (bottom panel). (C) APC-conjugated antibodies to cell-specific markers were used to determine percentages of blue cells for each population. (D) The percentage of each cell type in the total population of blue cells was determined by gating on each cell-specific marker. (E) Total population of each cell type in infected spleens.
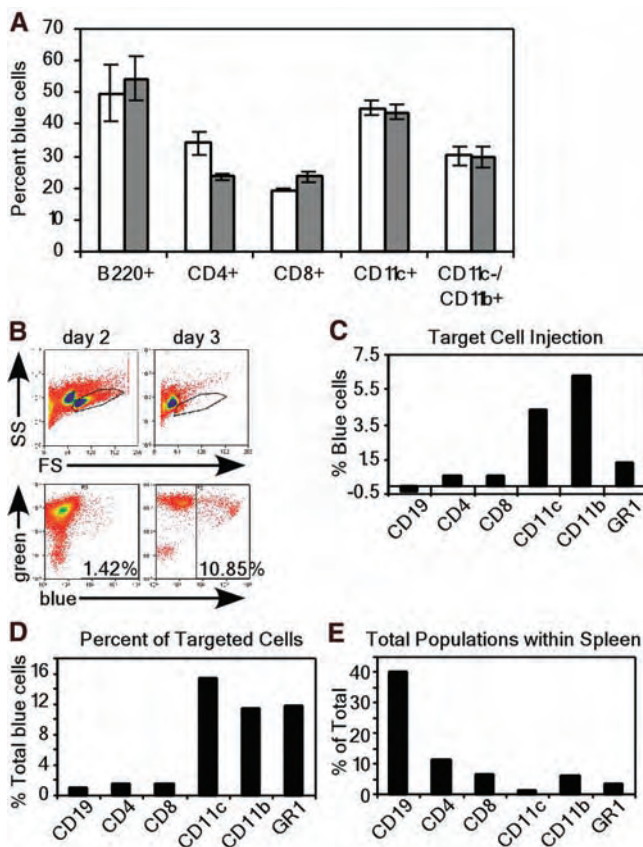
cent cells. *Y. pestis* injected all five cell types with YopM-Bla or YopE-Bla (Fig. 4A). After in vitro infection, injection of YopM-Bla and YopE-Bla occurred more frequently in B cells and dendritic cells, whereas CD8$^+$ T cells and macrophages were injected less frequently. Surprisingly, YopM-Bla was injected more frequently than YopE-Bla into CD4$^+$ T cells, whereas all other cell types were injected with both type III effectors at similar rates. Thus, there appear to be differences in the in vitro selection of immune cells for type III injection by *Y. pestis*, and not all effectors may be injected with similar efficiency into specific cell types.

Although *Y. pestis* displays a preference for B cells and dendritic cells in vitro, true target cell selection cannot be measured under these conditions because microbes are not presented with a choice between different host cells. To address this problem, we examined target cell selection in vivo by infecting C57BL/6 mice with 10^3 colony-forming units (CFU) of *Y. pestis* strain KIM D27 carrying pMM83 or pHSG576. Moribund animals were killed 2 or 3 days after infection; spleens were removed, homogenized, and then stained with CCF2-AM. Stained splenocytes were subjected to flow cytometry, revealing a precipitous decline in live cells on the third day after infection (Fig. 4B). The number of KIM D27 (pMM83)–injected blue cells increased from 1.42% on day 2 to 10.85% on day 3. In contrast, KIM D27 (pHSG576) infection

generated only background staining of blue cells (0.80% and 1.76% on days 2 and 3, respectively). The 10-fold increase in blue cells between days 2 and 3 can be accounted for by an increase in bacterial load as well as a decrease in immune cells from the spleen (*20*).

To further analyze target cell injection, we labeled splenocytes with allophycocyanin (APC)–conjugated antibodies specific for CD19$^+$/CD11c$^-$ (B cells), CD4$^+$ (T helper cells), CD8$^+$ (T killer cells), GR-1$^+$ (granulocytes/neutrophils), and CD11c$^+$/CD11b$^-$ (dendritic cells) as well as CD11b$^+$/CD11c$^-$ (macrophages), and we quantified the number of cells positive for both the antibody and blue fluorescence (Fig. 4C). At 2 days after infection, significant amounts of macrophages, dendritic cells, and neutrophils fluoresced blue, indicating that these cells had been injected by *Y. pestis* with YopM-Bla. However, no significant type III injection was observed for CD4$^+$ and CD8$^+$ T cells or B cells (CD19$^+$). When compared with the total amount of YopM-Bla–injected (blue) cells, target selection occurred most frequently for dendritic cells, macrophages, and neutrophils, whereas B and T cells were not selected for type III injection (Fig. 4D). Blue staining of phagocytic cells during infection with *Y. pestis* KIM D27 (pMM83, expressing YopM-Bla) is most likely due to type III injection and not caused by phagocytosis of β-lactamase containing bacteria, because control infections with

Y. pestis KIM D27 (pMM91, expressing Gst-Bla) resulted only in background fluorescence of splenocytes (fig. S5). To obtain additional proof for this hypothesis, future studies will need to test its predictions on multiple levels of experimental specificity. Because of the marked decline in live cells isolated 3 days after infection, it was not possible to make a statistically valid analysis of type III–injected cell types. However, it seemed that the percentage of blue cells was similar for all cell types, which suggests that all immune cells eventually became injected. Because B cells and T cells—which constitute the overwhelming majority of splenic immune cells (Fig. 4E)—were not injected on day 2, selection mechanisms must exist for bacterial type III injection of target cells. Macrophages, dendritic cells, and granulocytes/neutrophils are early targets for injection. Thus, *Y. pestis* appears to make use of the type III secretion system to destroy cells with innate immune functions that represent the first line of defense, thereby preventing adaptive responses and precipitating the fatal outcome of plague.

References and Notes

1. G. R. Cornelis *et al.*, *Microbiol. Mol. Biol. Rev.* **62**, 1315 (1998).
2. G. R. Cornelis, *J. Cell Biol.* **158**, 401 (2002).
3. R. D. Perry, P. A. Harmon, W. S. Bowmer, S. C. Straley, *Infect. Immun.* **54**, 428 (1986).
4. R. Rosqvist, K.-E. Magnusson, H. Wolf-Watz, *EMBO J.* **13**, 964 (1994).
5. M.-P. Sory, A. Boland, I. Lambermont, G. R. Cornelis, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11998 (1995).
6. J. B. Day, F. Ferracci, G. V. Plano, *Mol. Microbiol.* **47**, 807 (2003).
7. V. T. Lee, D. M. Anderson, O. Schneewind, *Mol. Microbiol.* **28**, 593 (1998).
8. X. Charpentier, E. Oswald, *J. Bacteriol.* **186**, 5486 (2004).
9. G. Zlokarnik *et al.*, *Science* **279**, 84 (1998).
10. R. R. Brubaker, *J. Bacteriol.* **98**, 1404 (1969).
11. L. W. Cheng, D. M. Anderson, O. Schneewind, *Mol. Microbiol.* **24**, 757 (1997).
12. A. Allaoui, S. Woestyn, C. Sluiters, G. Cornelis, *J. Bacteriol.* **176**, 4534 (1994).
13. See supporting data on *Science* Online.
14. S. Håkansson, T. Bergman, J.-C. Vanooteghem, G. Cornelis, H. Wolf-Watz, *Infect. Immun.* **61**, 71 (1993).
15. S. Håkansson *et al.*, *EMBO J.* **15**, 5812 (1996).
16. V. T. Lee, C. Tam, O. Schneewind, *J. Biol. Chem.* **275**, 36869 (2000).
17. T. Une, R. R. Brubaker, *J. Immunol.* **133**, 2226 (1984).
18. G. W. Anderson Jr. *et al.*, *Infect. Immun.* **64**, 4580 (1996).
19. F. Sebbane, D. Gardner, D. Long, B. B. Gowen, B. J. Hinnebusch, *Am. J. Pathol.* **166**, 1427 (2005).
20. A. V. Philipovskiy *et al.*, *Infect. Immun.* **73**, 1532 (2005).
21. We thank V. Bindokas for assistance with fluorescence microscopy, J. Marvin for help with flow cytometry, W. Williams and N. Green for isolating primary immune cells, Z. Haga for the statistical analysis, and M. Casadaban for providing strains. We acknowledge membership in and support from the Great Lakes Regional Center of Excellence for Biodefense and Emerging Infectious Diseases Consortium (GLRCE) (National Institute of Allergy and Infectious Diseases award 1-U54-AI-057153).

NEW PRODUCTS

Mitochondria Isolation Kit

The Mitochondria Isolation Kit for Tissue is a new tool for isolation of intact mitrochondria from soft and hard tissue samples. The kit offers two methods for mitochondria isolation. The first method makes use of a patent-pending reagent-based procedure that enables simultaneous processing of multiple samples. The second method relies on traditional Dounce homogenization, with fewer steps, for tissue disruption and subsequent isolation of the organelle. Both procedures make use of differential centrifugation to separate the intact mitochondria using a benchtop microcentrifuge, with the procedures completed in less than 60 min. The mitochrondria can be used for downstream applications such as proteomics research, disease profiling, and metabolic studies.
Pierce For information 800-874-3723 www.piercenet.com

Cellular Analysis Software

The IN Cell Developer Toolbox is intuitive software that gives researchers the ability to customize image analysis routines and investigate multiple cellular events simultaneously. Developer Toolbox expands the suite of analysis modules for processing images generated by the IN Cell Analyzer 1000 or 3000 cellular imaging platforms. Systems biologists and drug discovery scientists in pharmaceutical, biotechnology, and academic research can use the flexible Developer Toolbox for cellular image analysis for performing basic research, screening, lead profiling, target validation, and assay development. The software can be used to interpret phenotypic changes of fixed or live cells for responses to potential drug therapies to establish target specificity, drug efficacy, or toxicity. By analyzing kinetic events or fixed endpoints, the IN Cell platform generates context-rich information necessary to understand the role genes and proteins play in disease.
GE Healthcare For information 732-457-8149 www.gehealthcare.com

BioSafe Centrifuge Rotors

New BioSafe AeroSeal covers fit Beckman Coulter's AllSpin JS-5.3 rotors. These rotors deliver a combination of speed and high throughput for a wide range of life science research applications and are compatible with the Avanti J-E and J-26XP centrifuges. The covers have been independently validated for microbiological containment at a third-party testing facility and are the most recent addition to Beckman Coulter's biosafety initiative, which offers multiple levels of biocontainment for laboratory safety.
Beckman Coulter For information 800-742-2345 www.beckmancoulter.com

> **For more information visit GetInfo,** *Science*'s new online product index at
> *http://science.labvelocity.com*
>
> From the pages of GetInfo, you can:
> - Quickly find and request free information on products and services found in the pages of *Science*.
> - Ask vendors to contact you with more information.
> - Link directly to vendors' Web sites.

Tissue Culture Inserts

ThinCert Tissue Culture Inserts provide permeable support for co-cultivation, cell migration, and transport process applications. The inserts feature polyethylene terephthalate (PET) membranes with well-defined uniform pore sizes that provide improved cell adhesion and assure optimal exchange of compounds and nutrients. ThinCert membranes are sealed, not glued, and will not curl after detachment. Self-positioning, hanging geometry, and eccentric insert

well positioning improve the pipetting window and microscopic visibility to both insert and multiwell plate compartments. A variety of membrane pore sizes are available in 6-, 12-, and 24-well formats with universal plate compatibility.
Greiner Bio-One For information 800-884-4703 www.gbo.com/bioscience

High Voltage Power Supplies

Developed with exclusive Low Stress (LS) technology, the LS Range of high voltage power supplies offers a new level of high voltage precision and stability for mass spectrometers. LS technology reduces noise from stressed module components. Building on the unique technology of the 10 Watt HP Series, the LS Range provides low noise and high stability for 1-watt applications. Ripple is reduced to 2 ppm, temperature coefficient is 10 ppm per degree Centigrade, and drift is 50 ppm throughout the operating range of <1kV to 50 kV.
Applied Kilovolts For information 410-308-4828 www.appliedkilovoltsms.com

Calorimeter

The Oxylet system is an open circuit indirect calorimeter. It measures the energy produced by an animal's respiration, which researchers can use to calculate oxygen consumption, carbon dioxide production, respiratory exchange rate, and energy expenditure. The Oxylet system can measure the following parameters: respiratory flow, tidal volume, breathing frequency, peak flow, thoracic gas volume, oxygen consumption, food intake, and more. The system includes a gas analyzer unit, an air supply unit and switching unit, a data acquisition unit, and metabolic chambers or treadmills (up to eight).
Stoelting For information 630-860-9700 www.stoeltingco.com/physio

Proteome Research Product

The first in a series of products for exploring the human proteome, PreserveXT-QML Polymeric Micelles will help researchers working with an important class of human proteins—membrane proteins. Membrane proteins represent 70% of known drug targets. PreserveXT-QML consists of a proprietary mixture of amphiphiles (a molecule that consists of at least two parts, one that is hydrophilic and one that is hydrophobic) designed to resemble the native membrane environment. It can solubilize membrane proteins and preserve their activity, even during extended incubations in multi-well microplate assays.
QBI Life Sciences For information 608-441-2950 www.qbilifesci.com